

FOREWORD

Opinions, interpretations, conclusions and recommendations are those of the author and are not necessarily endorsed by the U.S. Army.

____ Where copyrighted material is quoted, permission has been obtained to use such material.

____ Where material from documents designated for limited distribution is quoted, permission has been obtained to use the material.

____ Citations of commercial organizations and trade names in this report do not constitute an official Department of Army endorsement or approval of the products or services of these organizations.

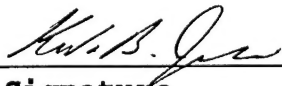
✓ In conducting research using animals, the investigator(s) adhered to the "Guide for the Care and Use of Laboratory Animals," prepared by the Committee on Care and use of Laboratory Animals of the Institute of Laboratory Resources, national Research Council (NIH Publication No. 86-23, Revised 1985).

✓ For the protection of human subjects, the investigator(s) adhered to policies of applicable Federal Law 45 CFR 46.

✓ In conducting research utilizing recombinant DNA technology, the investigator(s) adhered to current guidelines promulgated by the National Institutes of Health.

✓ In the conduct of research utilizing recombinant DNA, the investigator(s) adhered to the NIH Guidelines for Research Involving Recombinant DNA Molecules.

____ In the conduct of research involving hazardous organisms, the investigator(s) adhered to the CDC-NIH Guide for Biosafety in Microbiological and Biomedical Laboratories.


PI - Signature

8/8/99
Date

AD _____

Award Number: DAMD17-97-1-7097

TITLE: Functional Analysis of Nova Proteins: Gene Regulation and
Breast Tumor Immunity

PRINCIPAL INVESTIGATOR: Kirk B. Jensen, Ph.D.

CONTRACTING ORGANIZATION: The Rockefeller University
New York, New York 10021

REPORT DATE: July 1999

TYPE OF REPORT: Annual

PREPARED FOR: U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

DISTRIBUTION STATEMENT: Approved for Public Release;
Distribution Unlimited

The views, opinions and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy or decision unless so designated by other documentation.

DTIC QUALITY INSURED

20010124 093

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188), Washington, DC 20503.

1. AGENCY USE ONLY (Leave blank)		2. REPORT DATE July 1999	3. REPORT TYPE AND DATES COVERED Annual (1 Jul 98 - 30 Jun 99)	
4. TITLE AND SUBTITLE Functional Analysis of Nova Proteins: Gene Regulation and Breast Tumor Immunity			5. FUNDING NUMBERS DAMD17-97-1-7097	
6. AUTHOR(S) Jensen, Kirk B., Ph.D.				
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) The Rockefeller University New York, New York 10021			8. PERFORMING ORGANIZATION REPORT NUMBER	
9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES) U.S. Army Medical Research and Materiel Command Fort Detrick, Maryland 21702-5012			10. SPONSORING / MONITORING AGENCY REPORT NUMBER	
11. SUPPLEMENTARY NOTES This report contains colored photos				
12a. DISTRIBUTION / AVAILABILITY STATEMENT Approved for public release; distribution unlimited			12b. DISTRIBUTION CODE	
13. ABSTRACT (Maximum 200 words) The paraneoplastic neurological disorders (PNDs) are a group of autoimmune diseases in which neuron-specific proteins are expressed ectopically in tumors. Because PND patients have high-titer antibodies to these proteins, and display significant immunological suppression of tumors expressing these proteins, the PNDs are a model system in which to study tumor immunity. Paraneoplastic opsoclonus-myoclonus ataxia (POMA) is a PND in which the neuron-specific Nova-1 protein is expressed in tumors, particularly in breast cancer. Nova-1 is an RNA binding protein, and POMA patient antisera can block Nova-1's RNA binding activity. The action of POMA patient antisera suggests that the understanding of Nova-1's functional role in cellular RNA metabolism may provide insight into both breast tumor biology and in the elucidation of the effective anti-tumor immune response seen in POMA patients. We have determined the sequence-specific RNA binding activity of Nova-1, and show that the protein interacts with tetranucleotide RNA sequence UCAY. We have demonstrated that Nova-1 can function as an alternative splicing factor, which is dependent upon an intronic UCAY element. Furthermore, we have obtained a high-resolution crystal structure of the Nova-1 KH3 domain and a UCAY-containing RNA target. We have also generated Nova-1 null mice, which display specific alternative splicing defects as predicted by our <i>in vitro</i> studies. The Nova-1 null mice display severe motor dysfunction, suggestive of the human POMA disorder, and further characterization of these mice should aid in the elucidation of additional aspects of Nova-1 biology.				
14. SUBJECT TERMS Breast Cancer Nova-1, RNA binding protein, paraneoplastic neurological disorder, tumor immunity, gene regulation			15. NUMBER OF PAGES 69	
			16. PRICE CODE	
17. SECURITY CLASSIFICATION OF REPORT Unclassified	18. SECURITY CLASSIFICATION OF THIS PAGE Unclassified	19. SECURITY CLASSIFICATION OF ABSTRACT Unclassified	20. LIMITATION OF ABSTRACT Unlimited	

Table of Contents

Page 1.	Front Cover
Page 2.	Standard Form (SF) 298
Page 3.	Foreword
Page 4.	Table of Contents
Page 5.	Introduction
Pages 6-10.	Body
	Appendices

Introduction

The paraneoplastic neurological disorders (PNDs) are a group of autoimmune diseases in which neuron-specific proteins are expressed ectopically in tumors. Because PND patients have high-titer antibodies to these proteins, and display significant immunological suppression of tumors expressing these proteins, the PNDs are a model system in which to study tumor immunity. Paraneoplastic opsoclonus-myoclonus ataxia (POMA) is a PND in which the neuron-specific Nova-1 protein is expressed in tumors, particularly in breast cancer. Nova-1 is an RNA binding protein, and POMA patient antisera can block Nova-1's RNA binding activity. The action of POMA patient antisera suggests that the understanding of Nova-1's functional role in cellular RNA metabolism may provide insight into both breast tumor biology and in the elucidation of the effective anti-tumor immune response seen in POMA patients. We have determined the sequence-specific RNA binding activity of Nova-1, and show that the protein interacts with tetranucleotide RNA sequence UCAY. We have demonstrated that Nova-1 can function as an alternative splicing factor, which is dependent upon an intronic UCAY element. Furthermore, we have obtained a high-resolution crystal structure of the Nova-1 KH3 domain and a UCAY-containing RNA target. We have also generated Nova-1 null mice, which display specific alternative splicing defects as predicted by our *in vitro* studies. The Nova-1 null mice display severe motor dysfunction, suggestive of the human POMA disorder, and further characterization of these mice should aid in the elucidation of additional aspects of Nova-1 biology.

Body

Annual Summary

The investigation of the Nova-1 protein and its role in gene regulation has progressed substantially in the past reporting period. The Nova-1 protein is a neuron-specific RNA binding protein that has been found to be expressed in some in fallopian and breast tumors. Tumors that express Nova-1 can give rise to the paraneoplastic neurological disorder POMA (paraneoplastic opsoclonus myoclonus ataxia). This disorder is due to a robust immune response to the ectopically expressed Nova-1 in the tumor, which then is able to penetrate the blood-brain barrier and cause degeneration of Nova-1-expressing neurons. The POMA disorder, like all the paraneoplastic neurological disorders, provides a model system for the study of human tumor immunity, and the biological and functional study of the Nova-1 protein might illuminate a role for the Nova-1 protein in specific tumor gene regulation events.

The Nova-1 protein contains three KH-type RNA binding domains, and is similar to the ubiquitously expressed proteins hnRNP-E and hnRNP-K. These two RNA binding proteins have been implicated in several steps of RNA metabolism, including nuclear RNA export, RNA stability and translational control. Several other KH domain proteins have been shown to play a role in pre-mRNA splicing, including KSRP and SF1 in humans, Mer1p in yeast and Psi in *Drosophila*.

We have investigated Nova-1 from three different approaches: in vitro biochemical characterization of the protein, mouse genetic studies of Nova-1 function, and a biophysical approach to understand how Nova-1 interacts with RNA. Our in vitro biochemical studies began with identifying Nova-1's RNA targets. We have used an in vitro selection-amplification procedure known as SELEX (Systematic Evolution of Ligands by EXponential enrichment) to screen large ($\sim 10^{14}$) libraries of random RNA sequences for those with the highest affinity to Nova-1. The full length Nova-1 protein shows high affinity binding to the triplet repeat element (UCAUY)₃. We have used this sequence information to screen sequence databases for possible targets within the genome, and we have identified a candidate Nova-1 RNA target in the $\alpha 2$ inhibitory glycine receptor pre-mRNA present in an intron 85 nt upstream of the exon 3A. This exon is alternatively spliced in a mutually exclusive fashion with the downstream exon 3B. Immunoprecipitation of Nova-1 from brain extracts specifically co-precipitated GlyR $\alpha 2$ pre-mRNA, providing evidence that a Nova-1:GlyR $\alpha 2$ protein:RNA complex forms in neurons.

We have combined genetic and biochemical analyses of Nova-1 function to examine the hypothesis that Nova-1 regulates alternative splicing. We generated *Nova-1* null mice, which are born indistinguishable from their littermates, but die after birth with a profound motor failure that correlates with apoptotic death of motor neurons in the spinal cord and brainstem. Evaluation of GlyR $\alpha 2$ pre-mRNA splicing in *Nova-1* null mice reveals a specific defect in utilization of GlyR $\alpha 2$ E3A in the brainstem and spinal cord, but not in the forebrain, a region that does not express Nova-1. Moreover, a survey of neuronal exon usage in *Nova-1* null mice revealed an additional defect in inclusion of the alternatively spliced GABA_A $\gamma 2L$ cassette exon, which is preferentially used in the brainstem and spinal cord. The defect in GlyR $\alpha 2$ pre-mRNA alternative splicing in these mice is likely to be a direct consequence of the absence of Nova-1. Nova-1 in brain extracts specifically recognizes the GlyR $\alpha 2$ (UCAUY)₃ sequence motif, and Nova-1 acts directly on the (UCAUY)₃ element to upregulate utilization of GlyR $\alpha 2$ E3A in

cotransfection experiments. Taken together, these results demonstrate that Nova-1 is essential for postnatal motor neuron survival, where it functions to regulate neuron-specific splicing of inhibitory receptor pre-mRNAs.

Furthermore, to more precisely understand the molecular details of Nova-1's interaction with RNA, we undertook an additional SELEX experiment to find high affinity RNA targets for the isolated KH3 domain of Nova-1, with an eye of obtaining a Nova-RNA co-crystal structure. Using SELEX we demonstrated that the KH3 domain of Nova recognizes a single UCAY element in the context of a 20 base hairpin RNA; the UCAY tetranucleotide is optimally presented as a loop element of the hairpin scaffold, and requires protein residues C-terminal to the previously defined KH domain. These results suggest that KH domains in general recognize tetranucleotide motifs, and that biological RNA targets of KH domains may use either RNA secondary structure or repeated sequence elements to achieve high affinity and specificity of protein binding. This SELEX RNA ligand was then used to obtain a high-resolution co-crystal structure-- the first ever RNA complex with a KH-type RNA binding protein. Additionally, the Nova-1 KH3 SELEX results revealed a possible set of "rules" for the recognition of RNA by KH domain proteins. Specifically, the Nova-1 KH3 RNA target bears some striking similarities to known RNA targets of the KH domain proteins hnRNP-E and hnRNP-K, and suggests that Nova, like these two proteins, might play a role in cytoplasmic RNA metabolism, including translational control and RNA stability. This possibility, along with Nova's demonstrated role in alternative pre-mRNA splicing, suggests that Nova might indeed control some aspect of RNA metabolism when expressed ectopically in tumors.

Summary of Statement of Work Progress

Outlined below are the nine tasks encompassing the Statement of Work. The status of each task is given, followed by the task summary in italics, and concluding with specific comments on the reporting period's progress for each task.

Task 1: complete

RT-PCR of PND tumors samples for anonymous breast tumor specimens from POMA patients. Immunoreactive Nova also detectable in 10-20 % of non-PND breast tumors-- RT-PCR survey of non-PND anonymous tumor samples will be ongoing. Human breast tumor cell lines will be surveyed for Nova expression using POMA antisera reactivity and positive lines used for RT-PCR of Nova mRNAs.

Characterization of the Nova-1 protein has led to the identification of several alternatively spliced forms of the Nova-1 transcript. Our work on the Nova-1 null mouse has allowed us to completely characterize the exon/intron structure of the gene, giving us precise details on all possible forms of the Nova-1 gene product. The Nova-1 gene does indeed contain a single alternatively spliced exon, exon H, which codes for a 24 amino acid region immediately upstream of the second RNA binding domain. We believe that all other exons are constitutively spliced. Accordingly, we have constructed Nova-1 fusion protein vectors that express full-length Nova-1 including exon H, and we have just finished cloning Nova-1 fusion protein lacking exon

H. All of the POMA patient antisera tested in our laboratory is reactive with an epitope in the third RNA binding domain of Nova-1, which lies in the terminal exon of the gene. Thus, we believe that only full-length forms of Nova-1 are expressed in breast tumors of POMA patients.

Task 2: substantially complete

cDNA sequences from breast tumor forms of Nova will be placed in expression vectors to produce His-tagged fusion proteins for in vitro biochemical studies.

We have made expression constructs of both the exon H+ and exon H- forms of Nova-1 and are investigating the effect on this exon upon Nova's role in alternative splicing. Additionally, we are currently mapping the protein domains of Nova necessary for nuclear import and export of the protein.

Task 3: substantially complete

RNA selection using 52n RNA library and column chromatography. In vitro selection will be performed on purified Nova fusion protein from breast tumors or breast tumor cell lines. Selected RNA sequences will be cloned and sequenced for characterization and analysis.

RNA selection has been used to identify high-affinity RNA targets for both full-length Nova-1 and the Nova-1 KH3 domain. Current work seeks to identify high affinity RNA targets for the KH1 and KH2 domains of Nova-1, and we have obtained preliminary data that our "transcribed sequence SELEX" RNA target within domain IV of 28S rRNA does indeed bind KH1 and KH2.

Task 4: substantially complete

Coimmunoprecipitation from breast tumor specimen and mouse brain lysates of Nova protein and RNA bound by protein using POMA antibody

The task has been accomplished for mouse brain lysates, and has been used to show the in vivo interaction between Nova-1 and the $\alpha 2$ glycine receptor subunit pre-mRNA. Also, we have used coimmunoprecipitation of Nova-1 and RNA to construct our transcribed sequence SELEX libraries.

Task 5: complete

Sequencing of in vivo immunoprecipitation RNA, if possible. Construction of in vitro selection libraries from RNA recovered from in vivo immunoprecipitations. In vitro selection of libraries using tumor forms of Nova fusion protein.

The RNA obtained from task 4 has been used to construct a SELEX library of in vivo sequences that bind Nova-1. We have used this library in the transcribed sequence SELEX procedure to isolate and sequence two classes of high affinity sequences to Nova-1. The first set is the (UCAUY)₃ motif described above, and the second set is a ~150 bp sequence from domain IV of 28S rRNA.

Task 6: substantially complete

In vitro and in vivo selection ligands will be biochemically characterized: definition of ligand consensus sequences, mutational analysis of ligand sequences, and determination of protein:RNA binding constants and binding kinetics.

The (UCAUY)₃ motif RNA ligands, and the KH3 SELEX UCAC ligand have been extensively characterized. The 28S rRNA sequence that is necessary for high affinity binding to Nova-1 has been mapped to a core domain of approximately 110 bp. Determination of which KH domains of Nova-1 bind the 28S rRNA fragment are underway.

Task 7: complete

Definition of RNA-binding elements in breast tumor forms of Nova. Creation of truncated fusion proteins and point mutant variants.

The classification of all forms of Nova-1 protein, both in neurons and in tumors has been accomplished in tasks 1 and 2. The high affinity RNA targets identified by the in vitro SELEX methodology are also relevant to Nova-1's RNA targets in breast tumors.

Task 8: in progress

Genetic databank searching. Comparison of selected RNA sequences with databank matches, and determination of match relevance using mutational analysis results and coimmunoprecipitation experiments.

We have used our SELEX results of Nova-1 to search a database of intronic sequence adjacent to alternatively spliced exons, and identified the $\alpha 2$ glycine receptor subunit pre-mRNA. We are actively searching larger databases of mature mRNA sequence using the KH3 SELEX target sequence, since mutational and alignment data with know KH protein RNA targets suggests that Nova might play a role in cytoplasmic RNA metabolism binding to the untranslated regions of certain mRNAs.

Task 9: substantially complete

If experiments listed in technical objectives 2, 3 and 4 implicate Nova in splicing, then construction of relevant minigenes which preserve intronic and exonic structures necessary to obtain alternative splicing will be initiated. Minigene RNA splicing in will be assayed in the presence and

absence of breast tumor forms of Nova. If database searches implicate Nova in other areas of RNA processing, the appropriate assays will be developed and tested for Nova function. The functional significance of Nova gene regulation will be assessed in relation to breast tumor pathogenesis.

We have made considerable progress on functional studies of Nova-1's role in alternative splicing. Using an $\alpha 2$ glycine receptor subunit pre-mRNA minigene, we have shown in cotransfection studies that Nova-1 can bind to the intronic (UCAUY)₃ element adjacent to exon 3A, thereby upregulating utilization of this exon. This effect is consistent across all cell lines tested, and is abrogated by a point mutation in the (UCAUY)₃ element, demonstrating the specificity of the Nova-1:RNA interaction. Furthermore, we have demonstrated a specific defect in $\alpha 2$ glycine receptor exon 3A usage in the Nova-1 null mouse, and have identified a second splicing defect in the null mouse in the alternatively spliced GABA_A $\gamma 2L$ cassette exon. We are currently constructing a GABA_A $\gamma 2$ minigene to test whether we can duplicate the effects of the null mouse in cell transfection studies. If successful, we will map the binding site of Nova-1 within this minigene and show by mutational analysis that a Nova-1 binding site is necessary for proper regulation of GABA_A $\gamma 2L$ alternative splicing. The data gathered from this analysis will allow us to more precisely define, and search for, other candidate genes whose alternative splicing may be regulated by Nova-1. This will allow us to propose possible targets of Nova in breast tumors.

Appendices

Key research accomplishments

- Construction of Nova-1 null mice; elucidation of Nova-1 gene structure and all possible forms of Nova-1 protein.
- Expression and fusion protein constructs of the two forms of Nova-1: exon H+ and exon H- forms.
- In vitro selection of full-length Nova-1 and Nova-1 KH3 domain; identification of UCAY as core recognition element in RNA
- Identification by database searching of the $\alpha 2$ glycine receptor subunit pre-mRNA as a candidate Nova-1 target.
- Construction of a $\alpha 2$ glycine receptor minigene; demonstration that Nova-1 upregulates the use of exon 3A in cotransfection experiments.
- Demonstration that Nova-1 null mice show a specific splicing defect in exon 3A of the $\alpha 2$ glycine receptor; additional evidence that Nova-1 null mice show a defect in the use of the GABA_A $\gamma 2L$ cassette exon.
- High-resolution co-crystal structure of Nova-1 with a SELEX RNA target sequence; elucidation of how KH domain proteins recognize and discriminate RNA targets.
- Identification of possible “rules” for RNA recognition by KH domains; similarity seen between the Nova-1 KH3 SELEX consensus and known KH domain RNA targets.

Reportable outcomes

Manuscripts

Jensen, K.B., Musunuru, K Lewis, H.A, Burley, S.K., and Darnell, R.B. The tetranucleotide UCAY directs the specific recognition of RNA by the Nova KH3 domain. *PNAS*, *in press* (2000).

Jensen, K.B., Dredge, B.K., Stefani, G., Zhong, R., Buckanovich, R.J., Okano, H.J., Yang, Y.L., and Darnell, R.B. Nova-1 regulates neuron-specific alternative splicing and is essential for neuronal viability. *Neuron*, *in press* (2000).

Lewis, H.A, Musunuru, K, **Jensen, K.B.**, Edo, C., Chen, H., Darnell, R.B., and Burley, S.K. Sequence specific RNA binding by a Nova KH domain: Implications for paraneoplastic disease and the fragile X syndrome. *Cell* **100**, 323-332 (2000).

Invited Talks

“Functional analysis of the paraneoplastic neurological disease antigen Nova: An essential role for Nova in neuronal RNA processing.” The Department of Biochemistry and Molecular Biology, Division III, School of Chemistry. University of Barcelona, April 28, 1999.

“Identification of RNA sequences bound by the neuron-specific RNA binding protein Nova.” The Department of Biochemistry and Molecular Biology, Division IV, School of Pharmacy. University of Barcelona. April 26, 1999.

Poster Presentations

Dredge, B.K., **Jensen, K.B.** and Darnell, R.B. Alternatively regulated mRNAs in Nova-1 null mice. The RNA Society, Edinburgh, Scotland, June 1999.

Jensen, K.B., Dredge, B.K., Polydorides, A.D., Okano, H.J., and Darnell, R.B. Functional Analysis of the Onconeural Antigen Nova: Role of the Protein in Post-Transcriptional Gene Regulation. Keystone Symposium: The Molecular Basis of Cancer, Taos, MN, March 1999.

Poster Abstracts

Functional Analysis of the Onconeural Antigen Nova: Role of the Protein in Post-Transcriptional Gene Regulation.

Kirk B. Jensen, B. Kate Dredge, Alexandros D. Polydorides, Hirotaka J. Okano and Robert B. Darnell, The Rockefeller University, New York, NY 10021.

A rare group of neurological diseases, the paraneoplastic neurological disorders (PNDs), are autoimmune syndromes which appear to develop when tumors express proteins that are normally restricted to neurons. One of the best-studied PND antigens is Nova, an RNA binding protein of unknown function present only in tumor and normal brain tissues. The specific expression of Nova in tumors of gynecologic origin suggests that Nova's *in vivo* function maybe relevant to cancer biology. Moreover, the tumor suppression seen in patients with high antibody titers to Nova, and the demonstration that these antibodies block Nova binding to RNA suggest that tumor suppression and neuronal degeneration maybe linked to abrogation of Nova function.

The elucidation of Nova's biology requires the identification of those RNA sequences which are recognized by the protein, and an understanding of Nova's precise role in RNA processing events. We have identified several candidate messages with high affinity to Nova through both *in vitro* and *in vivo* biochemical approaches. We are using green fluorescent protein (GFP), as a reporter to assess Nova's involvement in a number of cellular mechanisms including alternative splicing, nuclear export and cytoplasmic processing of mRNA.

In addition to examining Nova function in cell lines, we have recently generated Nova-null mice. These animals undergo motor neuron apoptotic death and exhibit severe motor defects which resemble those seen in the human paraneoplastic syndrome. These mice provide a genetic system in which we are assessing Nova biochemical function *in vivo*.

Alternatively regulated mRNAs in Nova-1 null mice.

B. Kate Dredge, Kirk B. Jensen and Robert B. Darnell, The Rockefeller University, New York, NY 10021

The RNA binding protein Nova-1 was identified as the target antigen in a rare immune-mediated neurologic syndrome termed paraneoplastic opsoclonus myoclonus ataxia (POMA), a motor disorder affecting a subset of neurons in the brainstem, spinal cord and cerebellum. Nova-1 is a 55kDa protein harboring three KH domains whose expression is predominantly nuclear and strictly neuron-specific. This restricted expression pattern indicates that Nova-1 is involved in aspects of RNA processing which are specific for neuronal cells. Indeed, we have previously shown that Nova-1 binds to intronic sequence adjacent to alternatively spliced exons of the inhibitory glycine receptor $\alpha 2$ [1] and may regulate splice site selection.

We have generated Nova-1 null mice whose phenotype somewhat resembles that of POMA patients. The mice appear normal at birth, but become progressively debilitated over the

first postnatal week of life leading to death by P7, presumably attributable to motor weakness. Subsequently, we have been able to prolong their survival, up to approximately 21 days postnatal, by crossing them with CD-1 strain mice.

Using cDNA isolated from the CNS of Nova-1 null mice at two different time points we have screened commercial dot blots to identify mRNAs whose expression level is differentially regulated in knockout versus wild-type littermate mice. In addition we have performed SABRE (selective amplification via biotin and restriction mediated enrichment), a PCR-based selection procedure designed to specifically enrich differentially expressed mRNAs[2]. We are currently analyzing a number of interesting candidate genes cloned by this method using techniques which have been used successfully to analyze GlyR α 2, including RT-PCR. Preliminary data suggest changes in the expression levels of a number of apoptosis-related genes as well as several unknown genes. Further analysis will contribute to the elucidation of the role of Nova-1 in neuron-specific RNA processing and its implication in disease.

**The tetranucleotide UCAY directs the specific recognition of RNA by
the Nova KH3 domain.**

Kirk B. Jensen*, Kiran Musunuru* †, Hal A Lewis†, Stephen K. Burley ††
and Robert B. Darnell*

*Laboratories of Molecular Neuro-Oncology and †Molecular Biophysics,
‡Howard Hughes Medical Institute,
The Rockefeller University, 1230 York Avenue,
New York, NY 10021 USA

Running title: Nova KH3 binds UCAY

Classification: Biological Sciences, Biochemistry

Correspondence should be addressed to Dr. Robert B. Darnell, Laboratory of Molecular
Neuro-Oncology, The Rockefeller University, 1230 York Avenue, New York, NY 10021,
tel. 212-327-7460, fax 212-327-7147, email: darnelr@rockvax.rockefeller.edu.

Abstract

The Nova family of proteins are target antigens in the autoimmune disorder paraneoplastic opsoclonus-myoclonus ataxia (POMA), and contain KH-type RNA binding domains. The Nova-1 protein has recently been shown to regulate alternative splicing of the $\alpha 2$ glycine receptor subunit pre-mRNA by binding to an intronic element containing repeats of the tetranucleotide UCAU. Here, we have used selection-amplification to demonstrate that the KH3 domain of Nova recognizes a single UCAY element in the context of a 20 base hairpin RNA; the UCAY tetranucleotide is optimally presented as a loop element of the hairpin scaffold, and requires protein residues C-terminal to the previously defined KH domain. These results suggest that KH domains in general recognize tetranucleotide motifs, and that biological RNA targets of KH domains may use either RNA secondary structure or repeated sequence elements to achieve high affinity and specificity of protein binding.

Introduction

In eukaryotes, transcription of pre-mRNA is followed by a series of processing steps including mRNA splicing, polyadenylation, export, localization, translation and degradation. RNA-protein interactions are central in all of these pathways, and RNA binding proteins (RBPs) are often key regulators of posttranscriptional gene expression. The elucidation of how RBPs interact with RNA is thus critical to the understanding of these aspects of mRNA processing.

One of the most common protein motifs involved in RNA binding is the K-homology (KH) domain, originally described in the protein hnRNP-K (1). The KH domain consists of approximately 70 amino acids, and includes a conserved hydrophobic core, an invariant Gly-X-X-Gly motif and an additional variable segment; isolated KH domains of approximately this length are competent to bind ribohomopolymers (2-4). Over 50 KH domain-containing proteins have been identified; notably, many of these proteins contain multiple KH domains and/or other known RNA binding domains, such as the RNA recognition motif (RRM). NMR structural studies of individual KH domains revealed a conserved $\beta\alpha\alpha\beta\beta\alpha$ fold (3, 5), and high-resolution X-ray structures have additionally permitted the visualization of the Gly-X-X-Gly and variable loops (6).

How KH domain proteins function requires understanding how they interact with RNA, and the roles these interactions play in the context of a presumably complex and multicomponent regulatory system. The mammalian KH domain proteins KSRP and SF1, the yeast KH protein MER-1 and the *Drosophila* KH domain protein PSI have

all been implicated in regulating mRNA splicing (7-10). The heterogeneous nuclear RNP proteins hnRNP K and hnRNP E have shown to play important roles in mRNA stabilization and mRNA translational control (11-13). The proteins ZBP-1 and Vera, which each contain two RRM domains and four KH domains, have been shown to play a role in localizing particular mRNAs within the cell cytoplasm (14, 15). The absence of the mammalian KH domain protein FMR-1 is associated with the fragile-x mental retardation disorder, and a single amino acid mutation within its second KH domain impairs RNA binding and leads to a particularly severe form of mental retardation in humans (16-18). The functional roles of several of these KH domain proteins have been linked to their recognition of pyrimidine-rich sequences within their target RNAs, but we have yet to understand in detail how the specificity of RNA recognition is achieved and what other protein factors, if any, are necessary for KH domain protein function.

The Nova-1 and Nova-2 proteins (4, 19, 20) each contain three KH domains and are closely related to the hnRNP-K and hnRNP-E proteins (21). Nova proteins are exclusively expressed in CNS neurons, and are the target antigens in the autoimmune neurological disorder paraneoplastic opsoclonus myoclonus ataxia (POMA). We have previously used RNA selection-amplification (SELEX) (22, 23) to identify RNA stem loops containing the element (UCAUY)₃ as high affinity targets of Nova-1 (24), and have shown that Nova-1 specifically interacts with an intronic element of the same sequence in the pre-mRNA of the $\alpha 2$ glycine receptor subunit (24). Using both genetic and biochemical means, we have demonstrated that Nova-1 binding to this intronic site in the $\alpha 2$ glycine receptor can control the relative utilization of the two downstream,

alternatively spliced exons 3A and 3B (25). Additionally, selection-amplification experiments with Nova-2 revealed a consensus of GAGUCAU in a stem loop as a high-affinity target of the protein (20).

High-resolution X-ray structures of the third KH domain from Nova-1 and Nova-2 have been solved (6). This particular KH domain of Nova was chosen for structural study because it has been shown to be both necessary and sufficient for binding of the (UCAUY)₃ RNA motif (24). For the purposes of obtaining a Nova KH3/RNA co-crystal we have again used RNA selection-amplification to determine a high-affinity RNA target for the isolated Nova KH3 domain, and we report here the results of that screen. The RNA molecules selected using Nova KH3 contain a single UCAY element that lies within the loop of a 20-base hairpin structure. We have performed extensive mutagenesis on this molecule, confirming the strict requirement of the UCAY and its presentation within the context of a stem loop as necessary for recognition by Nova KH3. We have determined that several amino acids C-terminal to the previously defined KH domain are necessary for RNA binding, and we demonstrate that the specificity and affinity of this RNA:protein interaction is preserved in the full-length Nova protein. These results have led to the co-crystallization of Nova KH3 and one of the RNAs taken from this selection (26), and this structure shows excellent agreement with the mutagenesis and protein domain mapping studies reported here. Finally, we compare the Nova KH3 RNA ligand to known RNA sequences that interact with KH domain proteins and suggest that KH domains recognize RNA targets based upon a minimal four-base RNA sequence element. This interaction is weak, in both the

energetic and informational sense, and specificity for targets is likely achieved through the use of multiple KH domains within the same protein, the presence of multi-protein complexes on the RNA target, or, as demonstrated in the case of the Nova KH3 RNA target, an entropically favorable presentation of the core tetranucleotide to the KH domain.

Materials and Methods

KH domain expression and purification. The KH3/C-terminal domain from human Nova-1 (amino acids 423-510) and human Nova-2 (amino acids 406-492) were expressed and purified following the method of Lewis et al (6). The smaller protein fragments were prepared in an identical fashion. Briefly, N-terminal His-tagged fusion proteins were expressed in *E. coli*, and the insoluble fraction of the cell pellet was renatured in 6M guanidine-HCl and applied to a cation-exchange column. The purified, bound fusion protein was renatured on the resin, eluted and the His-tag cleaved using TEV protease (Gibco BRL). A final cation exchange column was used to isolate the cleaved KH3 domain. The proteins were assayed by electrospray mass spectrometry for correct molecular weight, circular dichroism for correct protein folding, and dynamic light scattering for monodispersity.

Selection-amplification. The synthetic oligonucleotide template 5'-TCCCGCTCGTCGTCT [25N] CCGCATCGTCCTCCCT-3', where 'N' indicates random incorporation of all four nucleotides, was prepared for 1st round transcription using Klenow fragment (Amersham/Pharmacia) and the oligonucleotide primer 5GL, 5'-GAAATTAATACGACTCACTATAGGGAGGACGATGCGG-3'. RNA was transcribed using T7 polymerase (Stratagene) to yield approximately 8 nmoles of full-length product (and approximately 10^{14} unique RNA molecules) for the first round of selection. Selection-amplification was performed essentially as described in Tuerk and Gold (22). Binding reactions were carried out using the buffer 1XBB (200 mM KOAc, 50 mM Tris-OAc pH 7.7 and 5 mM Mg(OAc)₂), using RNA:protein molar ratios that

varied from 5:1 at the beginning of selection to 250:1 at the end of selection. Partitioning was performed with 0.45 μ m nitrocellulose filters (Millipore). Bound RNA was extracted from the filters, reverse transcribed using the oligonucleotide primer 3GL, 5'-TCCCGCTCGTCGTCTG-3', and AMV-RT (Promega). Amplification of the library for the following rounds was done under mildly mutagenic PCR conditions using primers 5GL and 3GL, 1 mM dNTPs and 7.5 mM Mg(OAc)₂. Double stranded PCR product was concentrated using Microcon spin columns for transcription. Individual molecules were isolated from the selected pools using standard cloning and sequencing techniques.

Measurement of RNA-protein binding. Binding dissociation constants were measured either by a nitrocellulose filter binding assay (27) or by gel shift assay. For filter binding, 50 μ l reactions containing 50-100 fmol of RNA internally labeled with ³²P and concentrations of Nova-1 or Nova-2 KH3 in 3-fold dilutions typically ranging from 33 μ M to 45 nM were mixed in 1XBB and incubated at 10 min. at 25 ° C. Bound RNA was partitioned using nitrocellulose and the filters washed with 5 ml of 1XBB. The fraction of bound RNA was determined by scintillation counting. For gel shifts, binding reactions were modified as follows: 1XGS buffer (50 mM KOAC, 50 mM Tris-OAc, 10 mM DTT, 5 mM Mg(OAc)₂ and 30 μ g/ml tRNA) was substituted for 1XBB; reaction volume was 20 μ l; reaction temperature was 4° C; and Ficoll 400-DL (Sigma) was added to a final concentration of 2.5%. Samples were run on 8%, 37.5:1 acrylamide:bisacrylamide gels at 4° C at 100 V. Gels were dried for autoradiography and phosphorimaged (Molecular Dynamics) to quantitate bound and free RNA.

Dissociation constants were determined graphically by plotting the fraction of bound RNA versus the log of the protein concentration using KaleidaGraph software (28).

Transcription of oligonucleotide templates. The oligonucleotide for the 20-mer RNA 10021 was prepared using the synthetic oligonucleotide 5'-GCGGGGTGATCTTAGGTCCGCTATAGTGAGTCGTATTA-3'. Oligonucleotides for all 10021 mutants are based upon this sequence. The 10021 template was annealed to the oligonucleotide 5'-TAATACGACTCACTATAG-3' to provide a template for transcription by T7 polymerase, and RNA synthesis carried out using either α -³²P-UTP or -CTP in standard transcription buffer (Stratagene) supplemented with 80 mg/ml PEG 8000 (Sigma). Transcripts were size purified using 20% denaturing PAGE.

Results and Discussion

Selection-amplification of RNA for Nova KH3. The protein domains of Nova-1 (residues 423-510) and Nova-2 (residues 405-492) used for RNA selection-amplification include the KH3 domain and the remaining C-terminal extension of each Nova protein (the KH3-c constructs in Fig. 1A). These two Nova fragments are 81% identical, containing 7 conservative amino acid changes, and both proteins bound the round 0 pool of RNA with a K_d of greater than 5 μ M. While the previous X-ray structures of Nova lacked this C-terminal addition (Fig. 1A; KH3-a constructs) limited proteolysis of full length Nova-1 in the presence of a (UCAUY)₃ RNA indicated that this C-terminal protein fragment was protected by RNA binding, and thus possibly necessary for high affinity RNA binding to Nova KH3 (6). The Nova-1 and Nova-2 KH3 fragments were expressed as bacterial fusion proteins (see Materials and Methods), and the proteins purified to homogeneity by column chromatography.

The two KH3 protein fragments were assayed for binding to a (UCAUY)₃-containing 21-mer (mSB2), previously identified using selection-amplification with full-length Nova-1, and an identical RNA in which the core recognition sequence had been mutated to (UAAUY)₃ (mSB Δ 16) (24). While the isolated Nova KH3 domain retained a clear ability to bind and discriminate the wild-type UCAU-containing sequence from the mutant RNA, the overall K_d of the interaction was approximately 10 μ M (Fig. 1B), over 200-fold less than the K_d of the same RNA with full length Nova, and unsuitable for co-crystallization trials. To select an RNA with a higher affinity for the isolated

Nova KH3 domain, we employed a library with 25 random positions and performed selection-amplification individually on both Nova-1 KH3 and Nova-2 KH3 for 11 rounds. Figure 1B shows that the affinity of the round 0 pool for Nova-2 KH3 was detectable by filter binding at concentrations of protein approaching 10 μ M, but this interaction appeared quite non-specific as assayed by gel-shift assay (Fig. 1C).

The selected pools were cloned and sequenced at rounds 9, 10 and 11 and an aligned sequence set of the most abundant clones is displayed in Table 1. The largest sequence class (31% of the total), made up of sequences from both the Nova-1 and Nova-2 KH3 selections, is termed motif I. This set is characterized by a conserved bipartite sequence element GGACC[n3]WUCAYCCCC, where W is A or U, Y is U or C, and [n3] can be any three bases (29). The "phylogenetic" analysis of this motif's secondary structure suggests that these sequences form a 5-10 basepair stem-loop structure shown in Figure 2A. Clone 1-11-01, which contains both the conserved sequence element and a 10-base stem, binds the Nova KH3 domain with a K_d of approximately 500 nM by both filter binding and gel-shift assay (Figs. 1B and 1C). A second class of selected sequences, motif II, contains an almost identical conserved bipartite element GGAAC[n13]WUCAYCCCC, differing primarily in the distance between the two conserved elements. Motif II can be base-paired to give a stem-bulge-stem structure that is remarkably similar in secondary structure to that of motif I, with a second stem replacing the three variable positions of the motif I loop (Fig. 2A). The full length motif II clone 2-11-10 binds with a similar K_d to that of the motif I clones (data not shown). Another 22% of the cloned sequences contained a core UCAU or UCAC

element (sometimes multiple elements) but we have not assessed their secondary structures in detail (data not shown). An additional class appeared not to bind the Nova KH3 domain but show modest affinity for the selection-amplification partitioning matrix. Finally, seven orphan sequences could not be classified into any motif.

Nova KH3 binds a hairpin loop containing UCAY. We have made a detailed analysis of the interaction of the motif I type clones with the Nova KH3 domain. To determine the minimum RNA element necessary for specific binding to Nova KH3, we made a series of deletions to the putative 10-base stem of motif I clone 1-11-01. A 20-mer RNA, identical to 1-11-01 but with only a four-base stem, termed 10021 (Fig. 2B), was able to bind to both Nova-1 and Nova-2 KH3 with a K_d of 500 nM, identical to that of the full length parent clone (Fig. 2D). This core 20-base RNA was then used as a template for mutational studies. We tested 35 single point mutations in 10021, including all single base changes within the loop region of the conserved bipartite element ACC[n3]AUCACC. Furthermore, we also assayed several changes to the non-conserved loop region of the molecule and within the 4-base stem. All mutational analysis was carried out by gel-shift assay using Nova-2 KH3-c (Fig. 1), but identical results were obtained with Nova-1 KH3-c (data not shown).

The results of the mutational study are presented in Figure 2C. We have classified the binding characteristics of the mutants in three categories: 'tolerated,' with a small to non-detectable difference in binding to Nova KH3 (< 5fold); 'poor binders,' with significant change in affinity (> 50 fold); and 'non-binders,' in which there was no detectable binding at all at the highest concentration of protein used (> 33 μ M). Gel-

shift autoradiograms of 10021 and a representative mutant from each of the three categories of binding to Nova-2 KH3 is shown in Figures 2D. Within the first half of the bipartite ACC[n3]AUCACC element (A5-C7), a substitution of any of the three positions to a pyrimidine is well tolerated; purine substitutions either severely effect binding (A5→G; C7→A) or eliminate binding altogether (C6→G,A; C7→G). In the second half of the ACC[n3]AUCACC element (A11-C16), mutations as a whole are not tolerated. Within the 'core' UCACC sequence, only 2 out of 15 point mutations retain significant binding to Nova KH3. At three positions, C13, A14 and C16, all nine possible point mutants completely destroy the interaction between the RNA and the protein. At positions U12 and C15, pyrimidine substitutions are tolerated, but purines either cause severe binding impairment or complete loss of RNA binding. Thus, the UCACC 'core' is clearly, by mutational criteria, the most important region of the RNA for Nova KH3 recognition, with the C13-A14 dinucleotide being absolutely necessary for binding.

The stem nucleotides of 10021 also play an important role in binding to Nova KH3. The base pair most proximal to the loop, G4C17, was first assayed by reversing the bases to form the pair C4G17. This substitution bound Nova KH3 very poorly. Interestingly, the mutant G4U17, which replaces a G-C pair with a G-U pair, completely eliminated binding. Mismatches in this basepair result in poor binding (G4→A,C), or abolish binding altogether (C17→G). This asymmetry in mutation severity implicates C17 in a more important role in Nova KH3 recognition than G4, but clearly, basepairing of these two nucleotides is required for best binding. The identity of the preceding

basepair, G3C18, is also critical, with the reverse pair C3G18 severely affecting binding. However, the C18→U mutation, which substitutes a G-U pair for a G-C pair, does not significantly impair protein binding. The next base pair, C2G19, could be altered to either an A-U or U-A pair with no penalty on protein binding. The most distal base pair was not tested for mutations. We hypothesize that the contribution of the two distal base pairs is for purely structural (entropic) reasons, and that the stem as a whole is important in constraining the conformation of the loop and especially the UCAYC 'core'.

Finally, we find that point mutations within the variable U8-G10 region of the loop do not seem to affect Nova KH3 binding, although we have not exhaustively probed these positions. In contrast, deletion of two or three bases (Δ A9-G10 or Δ U8-G10) within this region abolishes protein binding, and even a single-base deletion (Δ G10) severely affects the protein-RNA interaction. Thus, it is likely that these bases function as a 'spacer'- providing the appropriate geometry of the loop for recognition by Nova KH3.

Full-length Nova also specifically recognizes the 10021 RNA. Previous data from our laboratory has demonstrated that the KH3 domain of Nova-1 is necessary and sufficient for binding to the sequence UCAY in RNA (24). To test whether the RNA ligands raised against Nova KH3 alone are capable of binding to full-length Nova protein, we performed filter binding assays with the complete Nova protein. The 10021 RNA does indeed bind full length Nova-1 with high affinity of approximately 8 nM as

shown in Figure 3A. The observed change in the dissociation constant seen with full length protein may be due to dimerization of the full-length Nova-1 protein in solution, as suggested by dynamic light scattering experiments (Lewis and Burley, unpublished observation); the presence of a second binding site for the 10021 RNA on the same Nova-1/Nova-1 complex would increase the apparent K_d of the interaction due to avidity effects. Alternatively, the RNA might in fact bind more strongly to the KH3 domain in the context of the entire protein. To test if the interaction with the 10021 RNA and full-length Nova-1 is specific, we also assayed the binding of the U12→A point mutant RNA, a 'poor binder' for Nova KH3. This RNA shows approximately 50-fold less affinity for the full-length protein than 10021 (Fig. 3A). Thus, the specificity of the Nova KH3 interaction for 10021 is intact even in full-length Nova-1. It seems unlikely that the higher affinity of the 10021 RNA for full-length Nova-1 is due to interaction with the KH1 and KH2 domains (see Fig. 3B and Conclusions).

RNA binding by Nova KH3 requires residues C-terminal to the KH domain.

Because our proteolysis protection data indicated that a UCAY RNA protected both regions of the KH3 domain and the adjacent C-terminus from digestion, we explored what fragment of Nova KH3 was necessary to retain binding to the 10021 RNA. Figure 3B displays gel shifts using the 10021 RNA with two Nova-2 KH3 constructs and the Nova-2 KH2 domain. The 10021 RNA binds the full C-terminal KH3 construct (KH3-c) with a K_d of approximately 500 nM. The dissociation constant using the N2 KH3-b construct (Fig. 1A), which is lacking the four terminal amino acids, binds 10021 identically (data not shown). However, when the N2 KH3-a construct is used, which

lacks the 15 most C-terminal amino acids, all binding to 10021 is lost (Fig. 3B). This is an important observation, since the Nova-2 KH3-a protein contains the entire KH domain as defined previously by protein sequence alignment (6). Thus, in the 11 amino acids that lie C-terminal to the defined KH domain are residues that directly or indirectly interact with the 10021 RNA, and are likely necessary for recognition of *in vivo* RNA targets of Nova. The Nova-2 KH2 domain (Fig. 3B) also completely fails to recognize the 10021 RNA.

The 10021 RNA as a general tetranucleotide scaffold for KH domain recognition. The Nova KH3 domain is a fully functional RNA binding domain, capable of a highly sequence specific interaction with RNA. Our selection-amplification experiments demonstrate that the core element of RNA recognition is dependent on the presence of the tetranucleotide UCAY, and this element is best presented to Nova KH3 in the context of a 12 base loop defined as ACC[n3]WUCAYC. An RNA target from this selection was recently co-crystallized with both the Nova-1 and Nova-2 KH3 domains ((6) and Lewis *et al.*, manuscript in preparation), and precisely confirms the conclusions of this present study. The crystal structure reveals the core UCAY of the RNA gripped in a 'molecular vise' formed by the cleft between the invariant and variable loops of the Nova KH3 domain. The vast majority of the RNA-protein contacts occur within the segment AUCAC (A11-C15), the RNA itself is in the predicted hairpin conformation, and there are additional stacking interactions continuing from the four-base stem through the next two nucleotide pairs, A5-C16 (which also forms one hydrogen bond), and C6-C15.

The selection-amplification, mutagenesis and co-crystal studies suggest that the minimal 20-base 10021 RNA hairpin is an optimal 'scaffolding' for the presentation of the core UCAY to Nova KH3, one that is likely to be reflected in biologically relevant Nova RNA targets. To examine the biological implications of the selection-amplification results we examined the published examples of RNA sequences that have been shown to interact with the KH domain proteins Nova-1, hnRNP-E, hnRNP-K, ZBP-1, and BBP/SF1. hnRNP-E1 and E2 are members of a protein complex (the α -complex) that bind to the 3' UTR of the α -globin gene and can stabilize the mRNA (13, 30). Similarly, a section of the 15-lipoxygenase 3' UTR, which contains a 19-base pyrimidine-rich repeated motif, can be crosslinked to hnRNP-E1 and K and is necessary and sufficient to confer translational silencing of a message by the two proteins (12, 31). Also, a 27-base sequence of the 3' UTR of β -actin has been shown to specifically crosslink to ZBP-1; this 'zip-code' sequence element is necessary and sufficient to confer localization of the β -actin message to the leading edge of fibroblasts. (15). The yeast branchpoint binding protein, BBP (32), and its mammalian orthologue SF1 (33) bind the conserved pre-mRNA branchpoint sequence UACUAAC (where the underlined A is the branchpoint) during commitment-complex formation in splicing.

Figure 4A displays the RNA targets for these KH domain proteins, along with the 10021 RNA. The boxed nucleotides in each figure indicate those bases critical for KH domain binding, and/or biological function. The β -actin RNA contains a GGACU element followed by ACACCC repeats (Fig. 4A). The GGACU-ACACCC elements are strikingly similar to the 10021 bipartite loop sequence, and the β -actin RNA can be base

paired to form a hairpin that would place these RNA elements in a manner spatially equivalent to the motif I and II sequences for Nova KH3 (Fig. 4B). Thus, perhaps the 10021 scaffolding may be employed in some biological targets of KH domain proteins. Despite the remarkable overall structural and sequence similarity between the zip element and 10021, Nova KH3 is able to absolutely discriminate between them (Fig. 4C). Therefore, the scaffolding may be an optimal strategy for the presentation of the core tetranucleotide to a KH domain, and we are currently exploring its affinity for the individual KH domains of ZPB-1. The RNA elements from $\alpha 2$ -glyR, α -globin, 15-lipoxygenase and the pre-mRNA branchpoint do not show any *obvious* secondary structures, but do share with 10021 a small pyrimidine/adenine element that is likely the core recognition element for their cognate KH domain proteins. A different strategy might be employed by $\alpha 2$ -glyR and 15-lipoxygenase targets, which repeat their core sequence elements from 3-10 times (see below).

Conclusions

An unexpected result of this study was the finding that the previously defined KH domain of Nova KH3 was not capable of high affinity binding of the 10021 RNA, requiring the addition of a number of amino acids C-terminal to the KH domain. The co-crystal structure revealed that A11 of 10021 interacts with an arginine residue seven amino acids C-terminal to the KH domain (6). A similar result was found for BBP, in which a construct retaining 31 amino acids C-terminal to the KH domain preserved specificity for the branchpoint sequence, while the KH domain by itself did not (34). Thus, at least for some KH domain proteins, residues proximal to the defined KH domain may be critical for proper RNA recognition.

The selection-amplification results for Nova KH3 agree well with previous SELEX experiments for the full length Nova-1 and Nova-2 proteins, in which UCAU was found as necessary for high affinity binding to the proteins (20, 24). The protein contacts made by the UCAY sequence in the Nova KH3 co-crystal cannot be duplicated by the Nova KH1 and KH2 domains (26), and biochemical data suggests that Nova KH3 is sufficient for binding of the full length Nova-1 SELEX targets (this study and (24)). Our preliminary selection-amplifications results suggest that Nova KH1 and KH2 recognize different RNA targets (K.B.J., K.M., and R.B.D., unpublished observations).

Finally, we have obtained data that implicates Nova-1 in the regulation of alternative splicing of the $\alpha 2$ glycine receptor subunit (25). The target recognized by Nova-1 in the glycine pre-mRNA is an intronic (UCAU)₃ element, and while it contains the absolutely conserved UCAY core necessary for Nova recognition, in secondary

structure it resembles more closely the in vivo targets of hnRNP-E and hnRNP-K than it does the 10021 hairpin of the in vitro selection experiments. It is likely that recognition of small RNA target sequences by KH domains in vivo involves several strategies. First might be the coordinate use of several KH domains within the same protein to select several small sequence elements in a longer RNA sequence. High affinity binding would be achieved only if several KH domains are able to bind to their individual core targets. Second, sequences such as the (UCAU)₃ repeat of the $\alpha 2$ glycine receptor subunit, or the 19 base element in the 15-lipoxygenase 3' UTR, might be recognized by a dimer or tetramer of a KH domain protein, which is rendered energetically favorable by multiple KH-RNA interactions, and by protein-protein contacts, as previously discussed for Nova-1 in this paper. Lastly, KH domain proteins might participate in protein-protein interactions with other RNA binding proteins, with recognition achieved coordinately by the two (or several) RNA binding proteins.

Acknowledgements

We thank B. Kate Dredge for critical reading of the manuscript and all members of our laboratories for help and many useful discussions. This work was supported in part by the NIH (R.B.D.). K.B.J. was supported by a Breast Cancer Research Program fellowship #DAMD 17-97-1-7097. K.M. was supported by the Tri-Institutional MD-PhD program and NIH MSTP grant GM07739.

Figures

Figure 1. **A.** Schematic representation of the Nova-1 and Nova-2 KH3 domain constructs used in this study. The boxed region indicates the KH domain (6). Amino acids that vary between Nova-1 and Nova-2 are underlined in the Nova-2 sequence. The arrows point to the C-terminal residue in the 'a', 'b' and 'c' forms of each protein. The longest KH3 forms, 'c', were used for the selection-amplification of RNA. **B.** Filter binding assays with Nova-2 KH3. The first graph displays the binding of a 21-mer RNA that contains a UCAUY triplet repeat (mSB2). The K_d of the RNA:protein interaction is approximately 10 μ M. The mutant RNA (mSB Δ 16), which contains a UAAUY triplet, shows no appreciable binding to Nova-2 KH3 even at 100 μ M protein. The second graph depicts the binding of the selection-amplification round 0 pool (pre-selection), with an apparent K_d greater than the highest concentration tested of 10 μ M. Also shown is the binding of a clone 1-11-01 isolated from round 11 of the selection. This clone has a K_d of approximately 500 nM. For clarity, all binding assays depicted employ Nova-2 KH3, although Nova-1 KH3 behaved in an identical fashion in all situations tested (data not shown). **C.** Gel shift assays using Nova-2 KH3. The round 0 pool does not exhibit any specific complex formation with Nova-2 KH3 at the highest protein concentration tested of 37 μ M. The selected clone 1-11-01 binds Nova-2 KH3 in a tight complex with a K_d of approximately 500 nM, similar to its dissociation constant as measured by filter binding assay.

Table 1. Results of selection amplification using Nova-1 and Nova-2 KH3 domains.

The first digit of each clone number indicate the clone's origin as from the Nova-1 or Nova-2 selection, the second two digits indicate the selection round number (9, 10 or 11), and the last two digits the individual clone number. Sequences are arranged in motifs according to sequence homology. Lowercase letters indicate the fixed sequence of the selection-amplification RNA. The core UCAY Nova KH3 recognition element is indicated by boldface in each sequence. Conserved nucleotides (including the two first basepairs of the conserved motif I and II stem) are underscored with dashes and potential basepairing interactions are indicated by a solid underscore. Sequences from individual clones which appear to be identical are represented only once with the number of individual isolates indicated in parentheses after the sequence.

Figure 2. A. Putative secondary structures of the motif I and II sequences. Uppercase letters represent the conserved nucleotides found in each sequence set. B. Predicted secondary structure of 10021, a 20-mer based upon the motif I sequence 1-11-01. 10021 binds with a K_d of 500 nM to Nova KH3, identical to the 1-11-01 parent clone. C. Table of the 10021 mutational analysis. The sequence of 10021 is displayed vertically at the left of the table. Point mutations at each nucleotide position are categorized into three classes according to their effect on the RNA's dissociation constant with Nova KH3: less than 5-fold decrease, greater than 50-fold decrease, and no binding at all to Nova KH3. Dissociation constants were measured by gel-shift assay. Listed at the bottom of the table are mutational changes to the 10021 stem and to the variable portion of the loop;

the effects on binding affinity are classified in accordance with the 10021 point mutants.

D. Autoradiograms of gel-shift assays with the 20-mer RNA 10021 and selected point mutants. Nova-2 KH3-c protein is titrated from left to right by three-fold dilutions starting at 37 μ M; additional conditions are described in the Material and Methods. A representative of each category of mutational defect is shown. A11→C is a tolerated mutation, with approximately 3-fold decrease in affinity for Nova KH3. A11→G is a poor binder, which shows a faint RNA:protein complex with a K_d approximately 100-fold greater Nova KH3 concentration than the wild-type 10021 RNA. Mutant C15→G is a non-binder, which does not form any complex with Nova KH3 even at the highest protein concentrations tested.

Figure 3 A. Nitrocellulose filter binding assay of the 10021 RNA and a point mutant RNA against full length Nova-1. The 10021 RNA binds full length Nova-1 with a K_d of approximately 5 nM. The point mutant 20-mer U12→A shows approximately 50-fold less binding to the full length Nova protein. **B.** Mapping the minimal protein domain of Nova KH3 necessary to bind the 10021 RNA. Protein is titrated from left to right by three-fold dilutions starting at 37 μ M. The top gel shift assay shows the binding of 10021 to the full C-terminal fragment of Nova-2 KH3 (N2 KH3-c). The K_d of the interaction is approximately 500 nM. The binding of 10021 to Nova-2 KH3-b (which lacks four C-terminal amino acids) is identical to the "c" construct. The middle gel-shift autoradiogram is the identical 10021 RNA assayed with the Nova-2 KH3-a protein, which is lacking the 15 C-terminal amino acids. There is no complex formation with

10021 and N2 KH3-a, demonstrating the necessity of some fraction of the 11 C-terminal amino acids in N2 KH3-b for RNA binding. The bottom autoradiogram is the same 10021 RNA assayed with the second KH domain of Nova-2 (N2 KH2). There is no complex formation with this related KH domain, demonstrating the specificity of the interaction.

Figure 4 A. Table of KH domain binding elements from β -actin (15), α 2-glyR pre-mRNA (24, 25), 15-lipoxygenase (12, 31), and α -globin 3' UTRs (35), and a canonical pre-mRNA branchpoint (34, 36). Uppercase letters represent the minimal sequence element necessary for interaction with the cognate KH domain protein. Boxed sequences indicate nucleotides most important for interaction with the protein, as determined by mutagenesis or deletion analysis. The UCAC sequence of 10021, which interacts directly with Nova KH3, is shown in red. Basepairing for 10021 is indicated by arrows; possible basepairing for the β -actin 3' UTR element is indicated by the dashed arrows.

B. Sequence and putative secondary structure of 10021 and the '10021-scaffold' RNA zip, which is based upon the β -actin RNA sequence. Changes to the 10021 molecule are indicated in blue; changes which are known to disrupt Nova KH3 binding are boxed in red.

C. Gel shift assay with 10021 and the zip 10021-scaffold RNA. Only 10021 forms a shifted complex with Nova KH3-c (protein concentration is 1 μ M).

References

1. Siomi, H., Matunis, M. J., Michael, W. M. & Dreyfuss, G. (1993) *Nucl Acids Res* **21**, 1193-1198.
2. Dejgaard, K. & Leffers, H. (1996) *Eur J Biochem* **241**, 425-31.
3. Musco, G., Stier, G., Joseph, C., Castiglione Morelli, M. A., Nilges, M., Gibson, T. J. & Pastore, A. (1996) *Cell* **85**, 237-45.
4. Buckanovich, R. J., Yang, Y. Y. & Darnell, R. B. (1996) *J Neurosci* **16**, 1114-22.
5. Baber, J. L., Libutti, D., Levens, D. & Tjandra, N. (1999) *J Mol Biol* **289**, 949-62.
6. Lewis, H. A., Chen, H., Edo, C., Buckanovich, R. J., Yang, Y. Y., Musunuru, K., Zhong, R., Darnell, R. B. & Burley, S. K. (1999) *Structure* **7**, 191-203.
7. Min, H., Turck, C. W., Nikolic, J. M. & Black, D. L. (1997) *Genes Dev* **11**, 1023-36.
8. Arning, S., Grüter, P., Bilbe, G. & Kramer, A. (1996) *RNA* **2**, 794-810.
9. Engebrecht, J. A., Voelkel-Meiman, K. & Roeder, G. S. (1991) *Cell* **66**, 1257-1268.
10. Siebel, C. W., Admon, A. & Rio, D. C. (1995) *Genes Dev* **9**, 269-283.
11. Gamarnik, A. V. & Andino, R. (1997) *Rna* **3**, 882-92.
12. Ostareck, D. H., Ostareck-Lederer, A., Wilm, M., Thiele, B. J., Mann, M. & Hentze, M. W. (1997) *Cell* **89**, 597-606.
13. Kiledjian, M., Wang, X. & Liebhaber, S. A. (1995) *Embo J* **14**, 4357-64.
14. Deshler, J. O., Highett, M. I., Abramson, T. & Schnapp, B. J. (1998) *Curr Biol* **8**, 489-96.
15. Ross, A. F., Oleynikov, Y., Kislauskis, E. H., Taneja, K. L. & Singer, R. H. (1997) *Mol Cell Biol* **17**, 2158-65.
16. Pieretti, M., Zhang, F., Fu, Y., Warren, S., Oostra, B., Caskey, C. & Nelson, D. (1991) *Cell* **66**, 817-822.
17. DeBouille, K., Verkerk, A., Reyniers, E., Vits, L., Hendrickx, J., Van Roy, B., Van Den Bos, F., de Graaff, E., Oostra, B. & Willems, P. (1993) *Nature Genet* **3**, 31-35.
18. Siomi, H., Choi, M., Siomi, M., Nussbaum, R. & Dreyfuss, G. (1994) *Cell* **77**, 33-39.
19. Buckanovich, R. J., Posner, J. B. & Darnell, R. B. (1993) *Neuron* **11**, 657-672.

20. Yang, Y. Y. L., Yin, G. L. & Darnell, R. B. (1998) *Proc. Natl Acad Sci* **95**, 13254-13259.
21. Ostareck-Lederer, A., Ostareck, D. H. & Hentze, M. W. (1998) *Trends Biochem Sci* **23**, 409-11.
22. Tuerk, C. & Gold, L. (1990) *Science* **249**, 505-510.
23. Ellington, A. & Szostak, J. (1990) *Nature* **346**, 818-822.
24. Buckanovich, R. J. & Darnell, R. B. (1997) *Mol Cell Biol* **17**, 3194-201.
25. Jensen, K. B., Dredge, B. K., Steffani, G., Zhong, R., Buckanovich, R. J., Okano, H. J., Yang, Y. Y. L. & Darnell, R. B. (1999) *Submitted*.
26. Lewis, H. A., Musunuru, K., Jensen, K. B., Edo, C., Chen, H., Darnell, R. B. & Burley, S. K. (2000) *Cell in the press*.
27. Carey, J., Cameron, V., de Haseth, P. L. & Uhlenbeck, O. C. (1983) *Biochemistry* **22**, 2601-10.
28. Irvine, D., Tuerk, C. & Gold, L. (1991) *J Mol Biol* **222**, 739-61.
29. Cornish-Bowden, A. (1985) *Nucleic Acids Res* **13**, 3021-30.
30. Wang, X., Kiledjian, M., Weiss, I. M. & Liebhaber, S. A. (1995) *Mol Cell Biol* **15**, 1769-77.
31. Ostareck-Lederer, A., Ostareck, D. H., Standart, N. & Thiele, B. J. (1994) *Embo J* **13**, 1476-81.
32. Abovich, N. & Rosbash, M. (1997) *Cell* **89**, 403-12.
33. Kramer, A. (1992) *Mol Cell Biol* **12**, 4545-52.
34. Berglund, J. A., Fleming, M. L. & Rosbash, M. (1998) *RNA* **4**, 998-1006.
35. Holcik, M. & Liebhaber, S. A. (1997) *Proc Natl Acad Sci U S A* **94**, 2410-4.
36. Berglund, J. A., Chua, K., Abovich, N., Reed, R. & Rosbash, M. (1997) *Cell* **89**, 781-7.

Table 1

Motif I

11003	gggaggacgaugcggACAGGACCCAGAU UCACCCC UGGCUGcagacgacgagcggga	
10904	gggaggacgaugcggUCAGGACCAACA UCACCCC UGUCCGcagacgacgagcggga	
10903	gggaggacgaugcggACCUAAA UCACCCC GCAUUACCCGCcagacgacgagcggga	
20902	gggaggacgaugcggUCAAGGAUCACGA UCACCCC UUGGCcagacgacgagcggga	
10901	gggaggacgaugcggUAGCAAGGACCUA UCACCCC UGcagacgacgagcggga	
10902	gggaggacgaugcggGGGGGACUGAU UCAUCCCC GCUGUGcagacgacgagcggga	
21112	gggaggacgaugcggAAAGANUGGGUAAU UCACCCC GCCGcagacgacgagcggga	
20903	gggaggacgaugcggAUUGCAUCACCA UCACCCC UCCCCcagacgacgagcggga	
11102	gggaggacgaugcggCUAACGACCAAA UCACACA UCGGCcagacgacgagcggga	(3)
20907	gggaggacgaugcggACCUAAAU UCACACC GCGAUGCCCcagacgacgagcggga	

Motif II

21110	gggaggacgaugcggAACGCGGAAGGUGCGCU UCACCCC Gcagacgacgagcggga	
21007	gggaggacgaugcggAACGCGGAG-GUGCGCU UCAUGCC UGcagacgacgagcggga	(3)
21107	gggaggacgaugcggUCAACCGUCCUUG---GCU UCAUACCCC cagacgacgagcggga	(2)
10906	gggaggacgaugcggACCA UCAUCAU --UAUA UCAUGCCC Gcagacgacgagcggga	

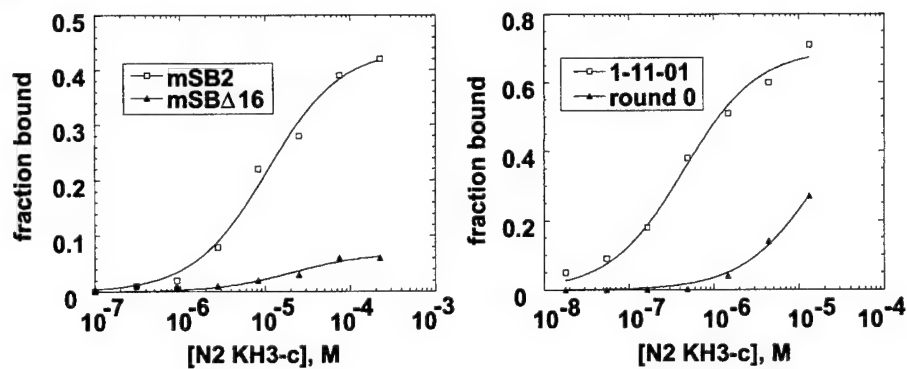
Figure 1

A

423 GAHMKDVVEI^aIAVPENLVGAILGKG^bGKTLVEYQELTGARIQISKKE
FVPGTRNRKVITITGTPAATQAAQYLITQRI^cTYEQGVRAANPQKVG **510**
Nova-1 KH3

405 GAHMKELVEI^aIAVPENLVGAILGKG^bGKTLVEYQELTGARIQISKKE
FLPGTRNRRVTITGSPAATQAAQYLISQ^cRVTYEQGVRAANPQKVG **492**
Nova-2 KH3

B



C

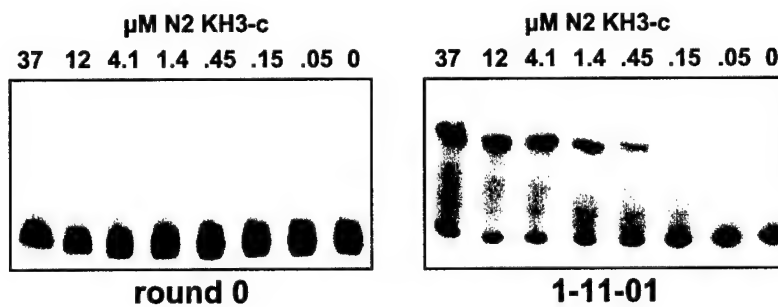
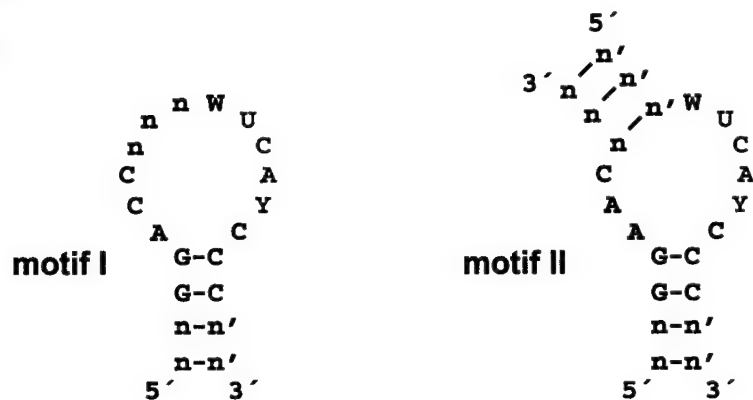
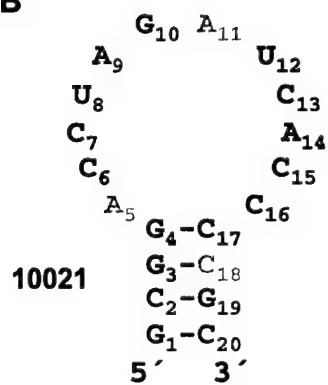


Figure 2

A



B



C

	<5 fold	>50 fold	no binding
G4		A, C	
A5	C, U	G	
C6	U		G, A
C7	U	A	G
U8	A		
A9	U		
G10	C		
A11	U, C		G
U12	C	G, A	
C13			G, U, A
A14			G, U, C
C15	U	A	G
C16			G, U, A
C17			G, U
C18	U		
C2G19	A-U, U-A		
G3C18		C-G	
G4C17			C-G

D

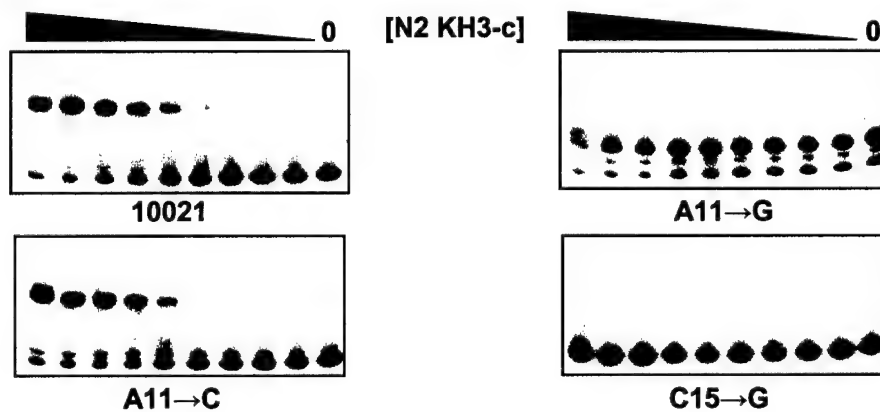


Figure 3

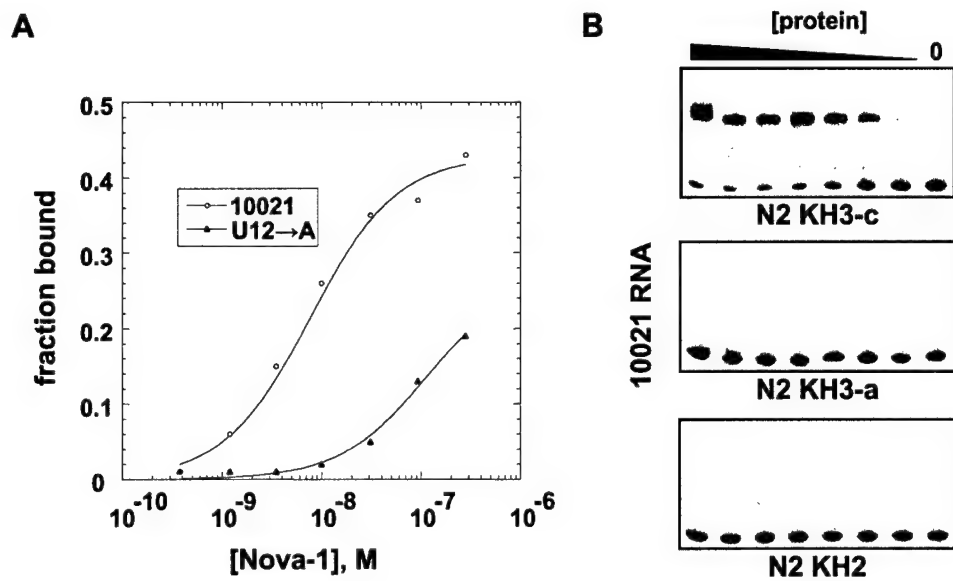
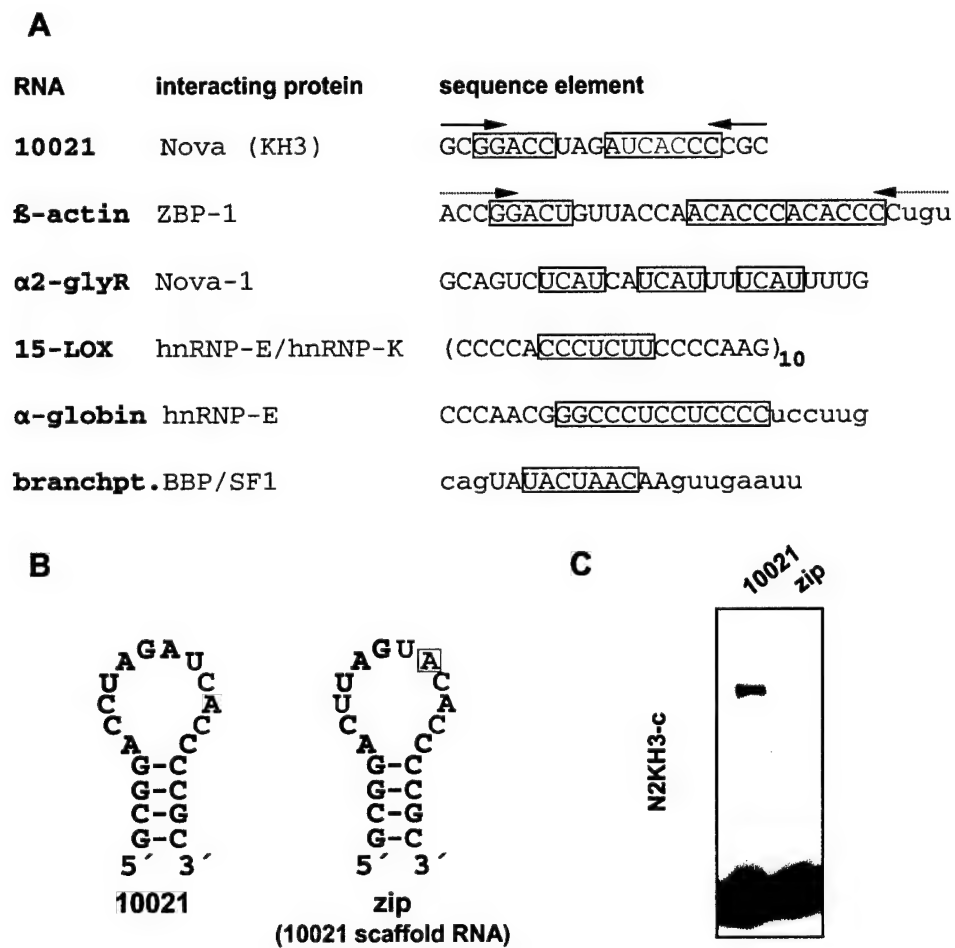


Figure 4



Nova-1 Regulates Neuron-Specific Alternative Splicing and Is Essential for Neuronal Viability

Kirk B. Jensen, B. Kate Dredge, Giovanni Stefani, Ru Zhong, Ronald J. Buckanovich, Hirotaka J. Okano, Yolanda Y. L. Yang, and Robert B. Darnell*
Laboratory of Molecular Neuro-Oncology
The Rockefeller University
New York, New York 10021

Summary

We have combined genetic and biochemical approaches to analyze the function of the RNA-binding protein Nova-1, the paraneoplastic opsoclonus-myoclonus ataxia (POMA) antigen. *Nova-1* null mice die postnatally from a motor deficit associated with apoptotic death of spinal and brainstem neurons. *Nova-1* null mice show specific splicing defects in two inhibitory receptor pre-mRNAs, glycine $\alpha 2$ exon 3A (GlyR $\alpha 2$ E3A) and GABA $_A$ exon $\gamma 2L$. Nova protein in brain extracts specifically bound to a previously identified GlyR $\alpha 2$ intronic (UCAUY)₃ Nova target sequence, and Nova-1 acted directly on this element to increase E3A splicing in cotransfection assays. We conclude that Nova-1 binds RNA in a sequence-specific manner to regulate neuronal pre-mRNA alternative splicing; the defect in splicing in *Nova-1* null mice provides a model for understanding the motor dysfunction in POMA.

Introduction

One of the fundamental mechanisms generating diversity in cells is differential processing of pre-mRNA. Alternative splicing can determine the subcellular location, molecular interactions, or function of proteins and is able to generate different sets of proteins from single RNA precursors. Alternative splicing is particularly evident in the nervous system, where it contributes to the complexity of neuronal function.

Proof of the importance of alternative splicing in cell development emerged from a combination of genetic and biochemical studies in *Drosophila*. Alternative splicing mediated by the Sxl protein is necessary and sufficient to regulate sex determination (reviewed by Baker, 1989; Tacke and Manley, 1999). Sxl is one of several *Drosophila* sequence-specific RNA-binding proteins that regulate splicing by either competing for or disrupting the binding of general splicing proteins, such as U2AF65 (Valcarcel et al., 1993). In contrast, Tra/Tra-2 act by attracting general splicing proteins to splice sites that are otherwise poorly utilized (Tian and Maniatis, 1992).

Alternative splicing in neurons was first demonstrated when a single pre-mRNA encoding calcitonin in thyroid cells was found to be differentially spliced to encode the neuropeptide neurotransmitter CGRP in neurons (Amara et al., 1982). Alternative splicing was subsequently found

to regulate a wide variety of neuronal activities, including the subcellular distribution and protein interactions of NR1 NMDA receptors (Ehlers et al., 1995, 1998); the physiology of glutamate (Sommer et al., 1990), NMDA (Hollmann et al., 1993), and GABA $_A$ (Macdonald, 1995) receptors; and the ability of agrin to induce clustering of acetylcholine receptors (Ferns et al., 1992). Alternative splicing of receptors is so widespread in the brain that it has been found in all major receptor subtypes, including the glycine, GABA (Macdonald, 1995), dopamine (Piccetti et al., 1997), serotonin (Claeysen et al., 1998), opioid (Zaki et al., 1996), and metabotropic glutamate (Conn and Pin, 1997) receptors, although the physiologic relevance of many of these splice variants is not yet certain. In some instances, alternative splicing has been found to generate a striking diversity of receptors. For example, alternative splicing generates over 1000 different receptor isoforms from three neurexin genes (Ullrich et al., 1995; Missler and Südhof, 1998) and 576 protein variants from the cSlo potassium channel gene in the avian cochlea (reviewed by Black, 1998), and may participate in regulating the utilization of variable exons in the neural cadherin-like genes to create a complex family of cell surface adhesion proteins (Kohmura et al., 1998; Wu and Maniatis, 1999). The greatest degree of functional diversity can be generated when splicing of individual exons is regulated as independent events—a single message that includes ten such alternatively spliced exons can generate 2¹⁰ or 1024 different protein products.

Our current understanding of how alternative splicing is regulated in mammalian cells derives almost exclusively from in vitro biochemical studies (Tacke and Manley, 1999), which have implicated sequence-specific RNA-binding proteins, including SR- and hnRNP-type proteins, in splicing regulation. Although these proteins are ubiquitously expressed, alternative splicing can be regulated in individual cells by varying their relative levels, as in the cases of the SR protein ASF-SF2 and hnRNP A1 (reviewed by Horowitz and Krainer, 1994), by regulating their posttranslational modifications (e.g., phosphorylation; see Petersen-Mahrt et al., 1999, and references therein) and by assembling the multiprotein complexes of the splicing apparatus (Hertel et al., 1997).

Ubiquitously expressed RNA-binding proteins have also been found to contribute to the regulation of neuron-specific splice variants. Neuron-specific splicing of calcitonin-CGRP pre-mRNA is regulated by the general factor polypyrimidine tract-binding protein (PTB), in addition to unidentified factors (Lou et al., 1999). Neurons insert a unique exon into the SH2 domain of c-src to generate the n-src protein, and this involves a number of generally expressed RNA-binding proteins, including PTB, hnRNP F, hnRNP H, and KSRP (Chan and Black, 1997; Min et al., 1997; Chou et al., 1999). Within the nervous system, the generation of GABA $_A$ $\gamma 2$ exon included (long, $\gamma 2L$) or excluded (short, $\gamma 2S$) forms is regulated by PTB and perhaps other proteins (Zhang et al., 1999). It remains unclear whether the regulation of tissue-specific splicing by general factors is able to account for the absolute differences in the ways in which

*To whom correspondence should be addressed (e-mail: darnell@rockvax.rockefeller.edu).

neurons and nonneuronal cells regulate the same mRNAs or whether they would be capable of generating the diversity seen in different types of neurons.

An interesting set of candidates for mammalian tissue-specific splicing regulators are neuron-specific RNA-binding proteins identified as paraneoplastic neurologic disease (PND) antigens (Darnell, 1996; Grabowski, 1998). The Nova family of proteins were identified as target antigens in the paraneoplastic opsoclonus-myoclonus ataxia (POMA) syndrome (Luque et al., 1991; Buckanovich et al., 1993; Yang et al., 1998), a disorder in which motor dysfunction suggests a defect in the inhibitory control of neurons in the brainstem and spinal cord. Nova proteins harbor three KH-type RNA-binding domains and are expressed exclusively in neurons within the central nervous system (CNS; Buckanovich et al., 1993, 1996). Nova proteins are closely related to hnRNP E1/E2 (Burd and Dreyfuss, 1994), which regulate α -globin stability (Ostareck-Lederer et al., 1998) and hnRNP K, which shuttles RNA transcripts from the nucleus to the cytoplasm (Michael et al., 1997), and both hnRNP K and hnRNP E1/E2 regulate lipoxigenase-15 mRNA translation in erythrocytes (Ostareck et al., 1997). Other KH-type RNA-binding proteins include the *Fragile-X* mental retardation gene product (Ashley et al., 1993; Siomi et al., 1993) and a number of KH domain proteins identified as splicing factors in yeast (MER-1; Engebrecht and Roeder, 1990), *Drosophila* (PSI; Siebel et al., 1995), and mammals (SF1, KSRP; Arning et al., 1996; Min et al., 1997).

RNA selection experiments demonstrated that Nova proteins are sequence-specific RNA-binding proteins (Buckanovich et al., 1996; Buckanovich and Darnell, 1997; Yang et al., 1998). Full-length Nova-1 binds with low-nanomolar affinity to a stem loop RNA harboring a sequence-specific motif (UCAUY)₃. This sequence led to the identification of a candidate Nova-1 RNA target in (GlyR α 2) pre-mRNA present in an intron 85 nt upstream of the inhibitory glycine receptor α 2 exon 3A (GlyR α 2 E3A; Buckanovich and Darnell, 1997). This exon is alternatively spliced in a mutually exclusive fashion with a downstream exon (E3B; Kuhse et al., 1991). Immunoprecipitation of Nova from brain extracts specifically coprecipitated GlyR α 2 pre-mRNA (Buckanovich and Darnell, 1997), providing evidence that a Nova-GlyR α 2 protein-RNA complex forms in neurons. Recent X-ray crystallography of a Nova-RNA complex confirms the specificity of the interaction between Nova and the core (UCAUY)₃ motif (Lewis et al., 2000).

In this report, we combine genetic and biochemical analyses of Nova-1 function to examine the hypothesis that Nova-1 regulates alternative splicing. We generated *Nova-1* null mice, which are born indistinguishable from their littermates but die after birth with a profound motor failure that correlates with apoptotic death of motor neurons in the spinal cord and brainstem. Evaluation of GlyR α 2 pre-mRNA splicing in *Nova-1* null mice reveals a specific defect in the utilization of GlyR α 2 E3A in the brainstem and spinal cord but not in the forebrain, a region that does not express Nova-1. Moreover, a survey of neuronal exon usage in *Nova-1* null mice revealed an additional defect in inclusion of the alternatively spliced GABA α γ 2L cassette exon, which is preferentially used in the brainstem and spinal cord (Zhang et al., 1996). The defect in GlyR α 2 pre-mRNA alternative splicing in

these mice is likely to be a direct consequence of the absence of Nova-1. Nova-1 in brain extracts specifically recognizes the GlyR α 2 (UCAUY)₃ sequence motif, and Nova-1 acts directly on the (UCAUY)₃ element to upregulate utilization of GlyR α 2 E3A in cotransfection experiments. Taken together, these results demonstrate that Nova-1 is essential for postnatal motor neuron survival, where it functions to regulate neuron-specific splicing of inhibitory receptor pre-mRNAs.

Results

Targeted Deletion of the Mouse *Nova-1* Gene

To assess the significance of Nova interactions with specific RNA ligands in neurons, we generated *Nova-1* null mice. We used homologous recombination to replace the coding sequence of each of the first two *Nova-1* exons, including the sequence encoding the initiator methionine, with a targeting vector encoding the neomycin resistance gene, which was flanked by loxP sites to allow Cre-mediated excision of the selectable markers (Figure 1A). In addition, the cassette harbored an IRES-*lacZ* sequence (Mombaerts et al., 1996) to allow visualization of *Nova-1*-expressing neurons. After transfection into embryonic stem (ES) cells, 2 of 96 G418-resistant clones had undergone homologous recombination (Figure 1B). These cells were transfected with a plasmid expressing Cre recombinase (pB5185) to excise the loxP-HSV-tk-neomycin (LTNL) cassette, reanalyzed by Southern blot analysis (data not shown), and used for blastocyst injection. Germline transmission was obtained in one line, and the genotype was confirmed by Southern blot (Figure 1C).

We analyzed Nova protein expression in wild-type and *Nova-1* null mice by Western blot analysis using POMA antisera, which recognize all Nova-related antigens. In the forebrain, a single band of 52 kDa and several larger bands are recognized (Figure 1D), all of which correspond precisely to the set of bands immunoprecipitated by a Nova-2-specific peptide antibody (Yang et al., 1998). In the hindbrain, additional bands of 52–55 kDa are evident in wild-type mice, and these bands are reduced and absent, respectively, in *Nova-1* heterozygous and null mice, indicating that they are derived from the *Nova-1* gene. This distribution of Nova-1 corresponds with previous immunohistochemical and in situ hybridization studies demonstrating that Nova-1 is specifically expressed in the hindbrain, spinal cord, and hypothalamus and is absent in the forebrain (Buckanovich et al., 1996; Yang et al., 1998). The distinct Nova-1 bands likely correspond to products of alternatively spliced *Nova-1* transcripts (Buckanovich et al., 1993) or to posttranslational modifications of the protein. We conclude that *Nova-1* null mice express no Nova-1 protein variants but are unchanged in their level of expression of Nova-2 and Nova-2-related proteins.

Nova-1 Is Required for Postnatal Survival of Neurons in the Brainstem and Spinal Cord

Nova-1 null mice were phenotypically indistinguishable from their littermates at birth. Thereafter, they demonstrated progressive motor dysfunction and died an average of 7–10 days after birth (Figure 1E). Prior to their

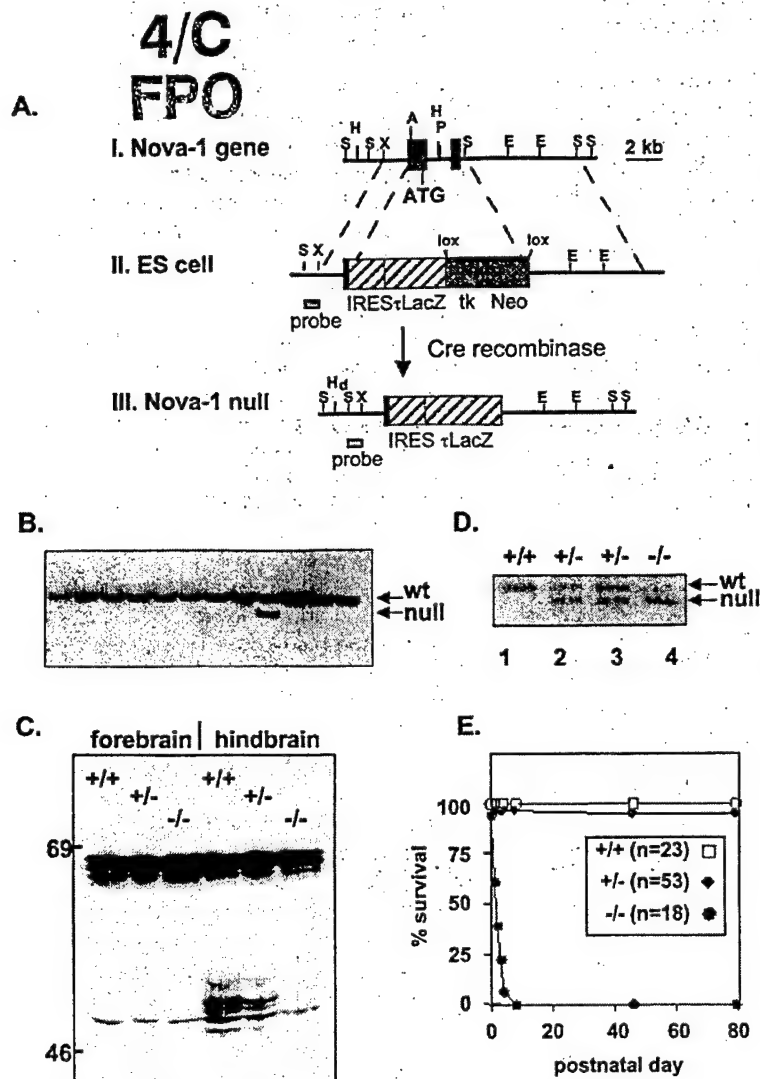


Figure 1. Generation of *Nova-1* Null Mice

(A) The wild type *Nova-1* locus illustrated contains the first and second coding exons (dark boxes, with initiator ATG indicated).

(AII) A targeting construct was generated harboring a genomic fragment (left: Xba-AlwNI, 1.6 kb) flanking the initiator methionine, an IRES-*tLacZ* LTNL insertion, and an intronic genomic fragment flanking the second coding exon (right: SacI-SacI 6kb).

(AIII) The *Nova-1* locus following Cre-mediated excision of the LTNL cassette. Restriction sites are indicated for SacI (S), HindIII (H), Xba (X), AlwNI (A), PacI (P), and EcoRI (E). The probe (S-X fragment) used for Southern blot analysis is indicated.

(B) Southern blot analysis of ES cells transfected with *Nova-1* targeting construct. The Southern blot of ES cell clones digested with EcoRI was probed with a DNA fragment that recognized a 7 kb band in the wild-type *Nova-1* locus and a 5 kb band in the recombinant *Nova-1* locus. The fidelity of homologous recombination was confirmed by probing the blot with a second probe flanking the right arm of the targeting construct (data not shown).

(C) Genotypic analysis of *Nova-1* null mice. Southern blot analysis was performed on tail DNA and digested with SacI, using the *Nova-1* genomic S-X probe described in (A). The genotypes for the samples are wild-type (+/+), *Nova-1* heterozygous (+/-), and *Nova-1* null (-/-).

(D) Western blot analysis of Nova protein expression. Extracts of mouse forebrain or hindbrain were made from wild-type, heterozygous, or *Nova-1* null mice; run on SDS-PAGE gels; and blotted with POMA antisera, which recognize all Nova protein species (Nova-1 and Nova-2; see Yang et al., 1998).

(E) Survival curve of *Nova-1* null mice and heterozygote and wild-type littermates. *Nova-1* null mice die on average between P7 and P10, whereas heterozygote mice are phenotypically normal and show no difference in survival from wild-type littermates. The ratio of genotypes is approximately 2:1:1 (het:wt:ko), suggesting there are no embryonic lethal effects in mice lacking *Nova-1*.

death, *Nova-1* null mice exhibited an action-induced tremulousness and overt motor weakness (Figure 2A), although they maintained normal responses to sensory stimuli. There was no abnormal phenotype in *Nova-1* heterozygotes.

Pathologic examination of newborn *Nova-1* null mice revealed normal gross structure and histology in the forebrain, brainstem, spinal cord, and extra-CNS structures. Histologic stains for β -galactosidase revealed weak β -galactosidase reactivity exclusively in CNS regions previously identified by in situ hybridization to express *Nova-1*, including the ventral spinal cord, brainstem nuclei, deep cerebellar nuclei, and hypothalamus (Buckanovich et al., 1996; Yang et al., 1998; Figure 2C). These results confirm that our targeting cassette was inserted accurately into the *Nova-1* locus and that *Nova-1* null neurons are confined to the subcortical structures we had previously identified.

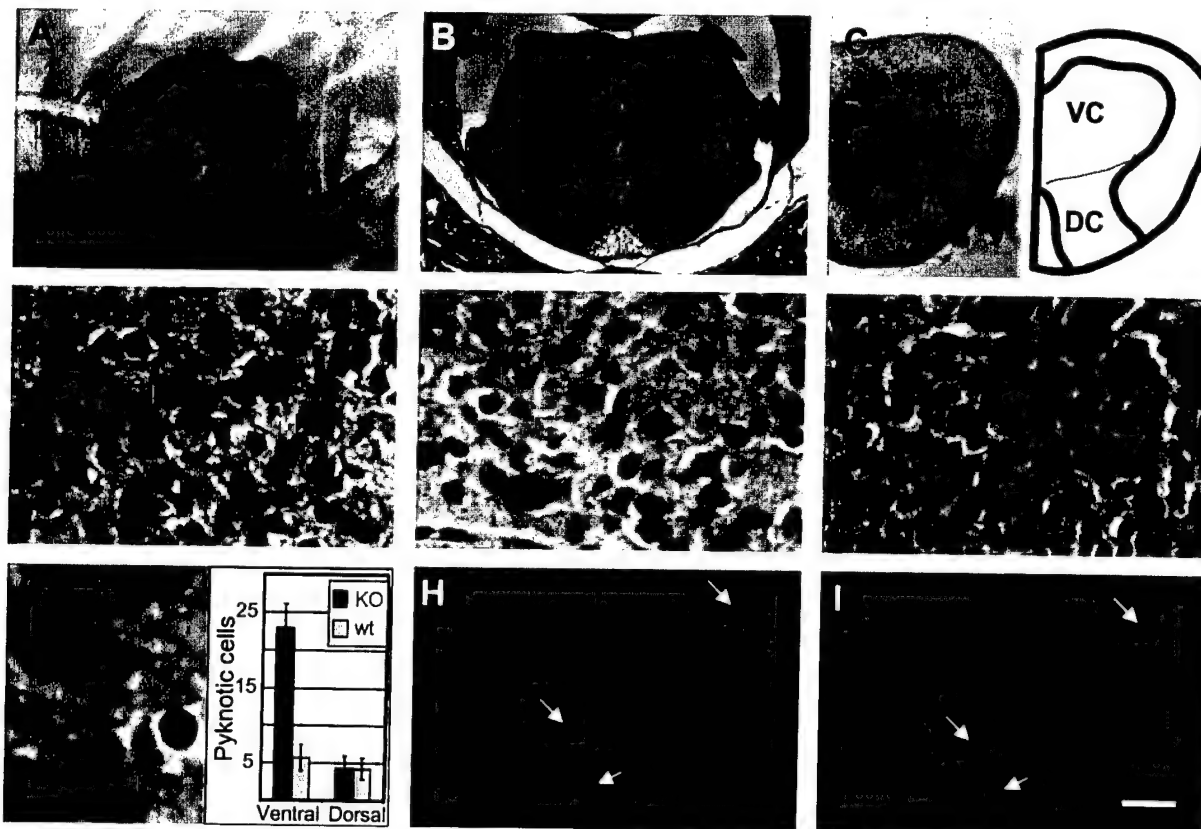
Histologic stains of older *Nova-1* null mice revealed a 4-fold increase in the number of pyknotic cells in the ventral region of the spinal cord relative to littermates (Figures 2D and 2E-2G). A survey of different areas revealed that the increase in pyknotic cells was most evident in the ventral spinal cord, in specific brainstem

nuclei, most notably in CN XII, and in the deep cerebellar nuclei but was not evident in regions that do not express *Nova-1*, including the dorsal spinal cord (Figures 2F and 2G), neocortex, and hippocampus (data not shown). TUNEL staining revealed that the pyknotic cells corresponded to cells undergoing apoptotic death (Figures 2H and 2I). Since Nova proteins are expressed exclusively in neurons, these results demonstrate that the hallmark pathologic change in postnatal *Nova-1* null mice is an increase in apoptotic death in hindbrain and ventral spinal cord neurons.

GlyR α 2 Splicing in *Nova-1* Null Mice

Since GlyR α 2 gene expression and physiology in rodent spinal cord and brainstem undergo regulation after the first week of postnatal life (Kuhse et al., 1991; Bechade et al., 1994; Singer et al., 1998), we bred *Nova-1* null mice into a CD1 genetic background, yielding *Nova-1* null mice that lived for an average of 2-3 weeks after birth. In these mice, the motor phenotype became particularly pronounced after ~7-10 days, with weakness and tremulousness, and with some mice showing clear signs of atrophy in the hindlimbs.

We examined GlyR α 2 splicing in postnatal day 16



NEURO1009.f2.4C

Figure 2. Histopathology of *Nova-1* Null Mice

(A) P5 *Nova-1* null (top) and wild-type littermate (below). *Nova-1* null mice are significantly smaller and were unable to walk without falling over at this age.

(B) Hematoxylin and eosin staining of a horizontal section of the spinal cord from P5 *Nova-1* null mouse. Boxed areas are shown under higher magnification in (D) and (dotted box) (F).

(C) Horizontal section of the spinal cord after X-gal staining of P5 *Nova-1* null mouse. Most of the neurons in the ventral motor column (VC) but not neurons in the dorsal sensory column (DC) were labeled blue.

(D–F) Solid boxed area in (B) is shown at high magnification in (D). Large numbers of pyknotic cells were observed in the ventral spinal cord of *Nova-1* null mouse (arrows) but not in the dorsal spinal cord (F) or the ventral spinal cord of a wild-type littermate (E).

(G) High magnification photomicrograph of a pyknotic cell in the ventral spinal cord of a *Nova-1* null mouse shows typical apoptotic morphology, including condensed staining of the nucleus. Inset shows quantitation of the number of pyknotic nuclei counted in the dorsal and the ventral spinal cords from 27 sections each of wild-type and *Nova-1* null spinal cords.

(H) TUNEL staining of the pyknotic cells (arrows) in ventral spinal cord of P5 *Nova-1* null mouse.

(I) DAPI counter staining of (H).

Scale bar, 125 μ m (B); 250 μ m (C); 31 μ m (D, E, and F); 10 μ m (G); and 20 μ m (H and I).

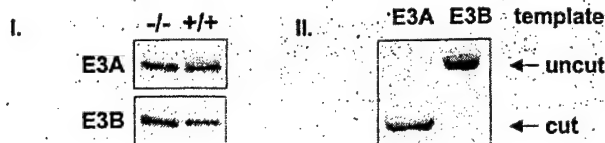
(P16) *Nova-1* null mice using several complementary methods. First, RNA was RT-PCR amplified using primers specific to GlyR α 2 E2 and GlyR α 2 E4. Reaction products were then digested with restriction enzymes that specifically recognized sites either in E3A (SspI) or E3B (BspMI). We confirmed that these enzymes uniquely recognized E3A or E3B (with >98% specificity; Figure 3AII) and digested E2-E4 PCR products with these enzymes to determine E3A and E3B splicing efficiency (Figure 3AI). In a second method, E3A- and E3B-specific primers were used under reaction conditions in which each primer showed absolute specificity for its respective exon (Figure 3B). In a third set of experiments, we performed RNase protection assays (see below). Quantitation of splicing using linear PCR reaction conditions gave highly reproducible results over a range of reaction conditions (input RNA, PCR cycle number) and between animals (see below), and demonstrated that the ratio of E3A/E3B splicing was decreased in *Nova-1* null mice relative to wild-type littermates by 2-fold.

To determine the reproducibility of these observations, we examined splicing to E3A and E3B in eight *Nova-1* null mice from six different litters and compared the results with splicing in normal littermates. In every case, RNA from *Nova-1* null mice showed a decrease in the utilization of E3A relative to E3B (Figure 3C). These results are consistent with our transfection data (see below), in which *Nova-1* acts to enhance the utilization of GlyR α 2 E3A, while in the absence of *Nova-1* in vivo there is a relative deficit in the ability of neurons to utilize the E3A splice site.

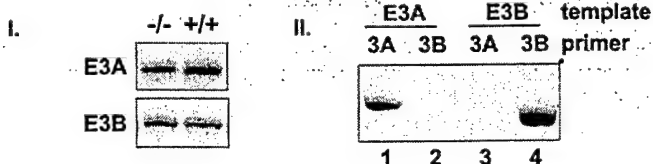
Specificity of Splicing Defects in *Nova-1* Null Mice

We next examined the specificity of the defect in GlyR α 2 splicing by comparing the splicing efficiency of six previously described alternatively spliced neuronal transcripts (Figure 4A). Splicing of five of these transcripts was unchanged in *Nova-1* null mice relative to their littermates, including clathrin light chain B (Clab), agrin,

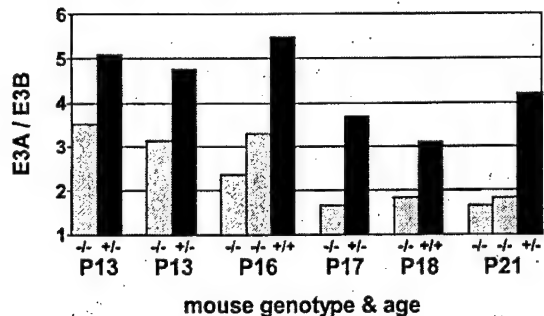
A. E2-E4 / Sspl



B. E3A/E3B primers



C.



isoforms. Data are presented as the ratio of E3A/E3B exon usage in each animal. PCR results were linear with respect to RNA input and cycle number. Actin RNA levels were measured in each sample by RT-PCR and differed by ~5%–10% between littermates (data not shown).

mGluR1, ICH-1, and n-src. However, alternative splicing of the GABA_A receptor $\gamma 2$ subunit was significantly altered in *Nova-1* null mice. The GABA_A $\gamma 2$ transcript is widely expressed in the brain, but the pre-mRNA is alternatively spliced to include a cassette exon, $\gamma 2L$, preferentially in brainstem and spinal neurons (Zhang et al., 1996). Examination of brainstem and spinal cord RNA revealed a 2-fold difference in the E3A/E3B GlyR $\alpha 2$ splicing ratio and nearly a 3-fold difference in the $\gamma 2L/\gamma 2S$ GABA_A $\gamma 2$ splicing ratio in wild-type versus *Nova-1* null mice, averaged over 11 and 9 litters, respectively (Figure 4B). Similar results were obtained with RNase protection assays (data not shown), although these results were less reliable than RT-PCR in quantitating the low-abundance GlyR $\alpha 2$ splice variants in a tissue sample of high complexity, as anticipated (Foley et al., 1993).

Although the GlyR $\alpha 2$ and GABA_A $\gamma 2$ genes are expressed throughout the brain, *Nova-1* expression is absent in rostral structures, including the forebrain and sensory thalamus (Buckanovich et al., 1996; Yang et al., 1998; Figure 1D). If aberrant GlyR $\alpha 2$ E3A and GABA_A $\gamma 2L$ splicing results directly from the lack of *Nova-1* in neurons, splicing of these RNAs should be normal in the forebrain and abnormal in the spinal cord of individual *Nova-1* null mice. Specific abnormalities in GlyR $\alpha 2$ and GABA_A $\gamma 2$ splicing were found in the spinal cord and hindbrain but not the forebrain (Figure 5; data not

shown). In forebrain of wild-type versus null mice, we found no differences in alternative splicing for GlyR $\alpha 2$, GABA_A $\gamma 2L$, or *Clab* mRNA or in steady-state levels of actin or the spinal motor neuron marker choline acetyltransferase (ChAT) (Figure 5). We conclude that *Nova-1* mediates a specific effect on the alternative splicing of GlyR $\alpha 2$ and GABA_A $\gamma 2$ receptor pre-mRNAs in a cell-autonomous manner.

Genotype and Developmental Profile of Splicing Changes

The level of *Nova-1* protein varies in a dose-dependent manner in wild-type mice relative to *Nova-1* heterozygous and null mice (Figure 1D). To assess whether the splicing defects of GlyR $\alpha 2$ and GABA_A $\gamma 2$ RNAs were dependent on the level of *Nova-1* protein, we examined the utilization of these exons in mice as a function of *Nova-1* gene dosage. The GABA_A $\gamma 2L$ splicing defect varied from 1.5-fold in heterozygous mice to a 3-fold change *Nova-1* null littermates (Figures 6A and 6B). GlyR $\alpha 2$ E3A utilization was only abnormal in *Nova-1* null mice, suggesting differences in the degree to which the *Nova-1* protein is rate limiting for proper splicing of GlyR $\alpha 2$ and GABA_A $\gamma 2$ pre-mRNAs.

When we examined alternative splicing as a function of age, we found that utilization of the GABA_A $\gamma 2L$ increased in wild-type mice during the first 3 weeks of

Figure 3. Aberrant GlyR $\alpha 2$ Splicing in *Nova-1* Null Mice

(A) RT-PCR analysis of GlyR $\alpha 2$ E3A and E3B splicing utilization. Spinal cord RNA from a P16 *Nova-1* null mouse and a wild-type littermate was analyzed by RT-PCR using ^{32}P -labeled E2 primer and unlabeled E4 primer followed by Sspl restriction digestion and autoradiography.

(B) Specificity of the restriction digest assay. PCR of a cDNA encoding the mature E2-E3A-E4 mRNA or the E2-E3B-E4 mRNA was digested with Sspl. Phosphorimager analysis of the data revealed that >98% of the E2-E3A-E4 DNA was digested, while <2% of the E2-E3B-E4 DNA was digested. Similar discrimination of E3A and E3B products was seen with BspMI digestions (data not shown).

(C) RT-PCR analysis of GlyR $\alpha 2$ E3A and E3B splicing utilization. Spinal cord RNA from a P18 *Nova-1* null mouse and a wild-type littermate was analyzed by RT-PCR using ^{32}P -labeled E2 primer and unlabeled E3A or E3B primer followed by autoradiography.

(D) Specificity of the assay. cDNA clones encoding the mature E2-E3A mRNA or the E2-E3B mRNA were PCR amplified with ^{32}P -labeled E2-E3A or E2-E3B primer sets. Analysis of this data on a phosphorimager revealed that E2-E3A or E2-E3B products could not be detected using the mismatched primer pair (E2-E3B or E2-E3A primers, respectively).

(E) Reproducibility of the GlyR $\alpha 2$ splicing defect in *Nova-1* null mice. Spinal cord RNA from *Nova-1* null mice and their littermates with indicated genotypes was analyzed as in (B) for the presence of GlyR $\alpha 2$ E3A and E3B

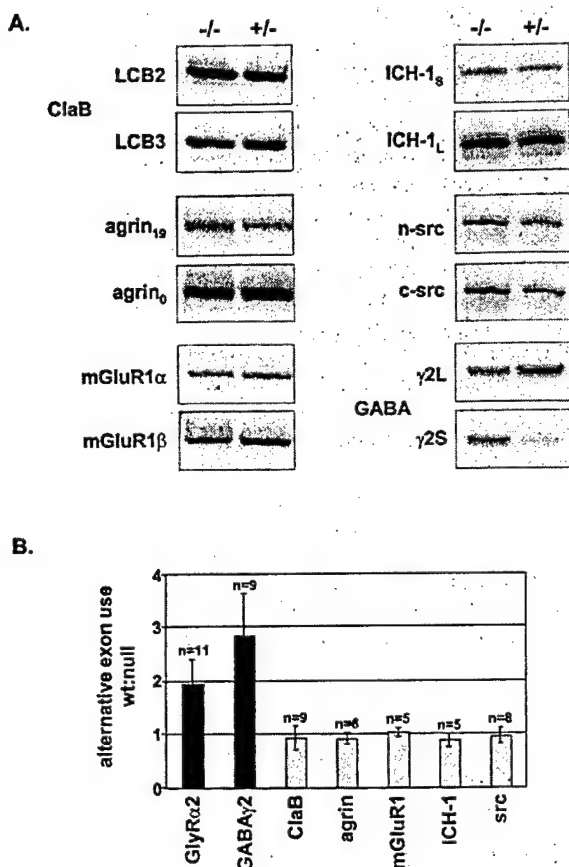


Figure 4. Specific Abnormalities in Alternative Splicing of GlyRα2 and GABA_A γ2 Transcripts in *Nova-1* Null Mice

(A) RNA isolated from a P16 *Nova-1* null mouse and a heterozygous littermate was analyzed by RT-PCR with primers specific to the indicated genes. All PCR reactions were performed in duplicate using minus-RT reactions (data not shown). LCB2 includes and LCB3 lacks the neuron-specific EN exon in the ClaB mRNA (Wang and Grabowski, 1996). Agrin₀ lacks and agrin₁₉ includes both Z site exons (Ferns et al., 1992). mGluR1α lacks and mGluR1β contains an additional exon in the metabotropic glutamate receptor mRNA (Hollmann and Heinemann, 1994). ICH-1S includes and ICH-1L lacks an additional exon in the IL-1β-converting enzyme mRNA (Wang et al., 1994). n-src mRNA includes a neuron-specific exon in the c-src message (Martinez et al., 1987).

(B) Quantitation of the data presented in (A), together with additional RNA samples (the number of litters examined is indicated by n). The ratio of long to short alternatively spliced products measured in wild-type animals relative to the respective ratio in the corresponding *Nova-1* null littermate is presented. For GlyRα2, the ratio of E3A/E3B in wild-type versus *Nova-1* null littermate is depicted. The bars represent the average change in wild-type:null exon usage for the number of litters examined; error bars represent the standard deviation.

development (Figure 6C), consistent with previous analyses (Wang and Grabowski, 1996; Zhang et al., 1996), and that a deficiency in exon utilization first becomes apparent at P7 in *Nova-1* null mice (Figure 6C). In *Nova-1* heterozygous mice, there was again an intermediate deficit in GABA_A γ2 usage evident at P7. The GlyRα2 E3A/E3B ratio was unchanged in P0–P4 *Nova-1* null mice (<10% difference across multiple littermates) and thereafter differed by 2-fold (data not shown).

Mouse Brain Nova Binds Specifically to (UCAUY)₃ RNA

To evaluate whether the GlyRα2 splicing defect evident in *Nova-1* null mice could be a direct action of Nova-1 on GlyRα2 pre-mRNA, we first examined whether native Nova protein from mouse brain is able to specifically bind the GlyRα2 intronic (UCAUY)₃ element. Previous work demonstrated that immunoprecipitates of Nova protein from brain extracts contain GlyRα2 pre-mRNA (Buckanovich and Darnell, 1997) but did not define the specific sequence to which Nova bound. To examine this question, a 90 nt intronic RNA surrounding the GlyRα2 (UCAUY)₃ element was transcribed in the presence of ³²P-UTP and ultraviolet (UV) cross-linked to proteins in brain extracts. As a control, we also transcribed a mutant GlyRα2 RNA, (UAAUY)₃, previously shown not to bind Nova-1 in vitro (Buckanovich and Darnell, 1997). These mixtures were RNase treated, and Nova protein was immunoprecipitated, run on SDS-PAGE gels, and transferred to nitrocellulose. A ³²P-labeled RNA band superimposable on the 55 kDs Nova protein band was seen specifically in immunoprecipitates of extracts cross-linked to the wild-type (UCAUY)₃ but not the mutant (UAAUY)₃ RNA (Figure 7). This result demonstrates that native neuronal Nova protein specifically binds to the GlyRα2 intronic (UCAUY)₃ RNA.

Nova Regulates GlyRα2 Pre-mRNA Alternative Splicing in Cell Lines

We next examined whether Nova-1 is able to directly act on GlyRα2 pre-mRNA to regulate alternative splicing of the E3A/E3B exons. We generated a minigene construct consisting of genomic DNA encoding GlyRα2 E2, E3A, E3B, and E4 and several hundred nucleotides of intronic DNA on either side of these exons (Figure 8A). To allow us to determine whether Nova-1 acts directly on the GlyRα2 intronic (UCAUY)₃ element, we also generated a mutant minigene in which this element was mutated to the nonbinding (UAAUY)₃ sequence. These constructs were then transfected into three different cell lines along with increasing amounts of a plasmid expressing Nova-1 (pcNova-1).

Nova-1 transfection mediated an increase in GlyRα2 E3A splicing in each of the three cell lines tested. Figure 8B illustrates the results of splicing assays performed on extracts of a neuroblastoma cell line, N2A, transfected with the wild-type (UCAU) or mutant (UAAU) GlyRα2 minigene in the presence or absence of pcNova-1. A Nova-1 dose-dependent increase in splicing to E3A (and compensatory decrease in E3B) is evident in cells transfected with the wild-type UCAU but not the mutant UAAU GlyRα2 minigene. We demonstrated that Nova-1 was present in increasing amounts in cells transfected with increasing amounts of pcNova-1 by Western blot analysis (Figure 8C).

To confirm these results and analyze splicing to both E3A and E3B, we repeated independent transfection experiments in N2A cells (Figure 8D). In cells transfected with the wild-type UCAU GlyRα2 minigene, Nova-1 reproducibly increased E3A splicing at the expense of E3B splicing. Nova-1 mediated a 2-fold increase in E3A utilization, an effect similar in magnitude to the decrease in the E3A/E3B splicing ratio seen in *Nova-1* null mice (Figures 3 and 4).

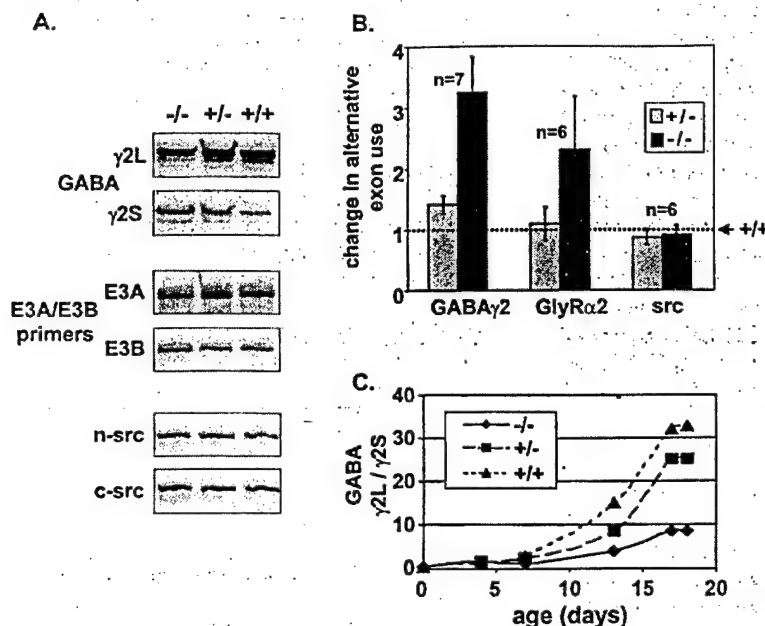


Figure 6. Changes in GlyR $\alpha 2$ and GABA $\gamma 2$ Alternative Splicing as a Function of Gene Dosage and Development

(A) GABA $\gamma 2$, GlyR $\alpha 2$, and c-src alternative exon usage in the spinal cords of a P16 *Nova-1* null mouse and its heterozygous and wild-type littermates was analyzed. There was a significant difference in alternative exon utilization in both the null and heterozygote mice for the GABA $\gamma 2$ message, whereas the GlyR $\alpha 2$ E3A/E3B ratio is only disturbed in the null mouse. Alternatively spliced n-src mRNA is unchanged in all three genotypes.

(B) Data obtained from (A) and the indicated number of additional litters are shown. GlyR $\alpha 2$, GABA $\gamma 2$, and n-src alternative exon usage in *Nova-1* null and heterozygote mice is plotted using the alternative exon ratio of the wild-type mouse set to 1. Bars represent the average of the normalized exon use ratio; error bars indicate the standard deviation.

(C) Developmental change in the GABA $\gamma 2L/\gamma 2S$ exon use ratio. Points represent single measurements at the indicated developmental times.

In N2A cells transfected with the mutant UAAU GlyR $\alpha 2$ minigene, splicing to E3A was unchanged relative to the wild-type minigene (Figure 8D). In the presence of increasing amounts of Nova, a paradoxical increase in E3B, at the expense of E3A, was seen. While the mechanism underlying this observation is unclear, a possible explanation is that the UAAU mutation allowed detection of a cryptic Nova-1 binding site able to mediate an action

on E3B splicing. Taken together, these results demonstrate that the action of Nova to enhance E3A utilization in wild-type UCAU GlyR $\alpha 2$ pre-mRNAs was abolished by mutation in the intronic UCAU Nova-binding element.

We repeated these experiments in HeLa cells and 293T cells, two nonneuronal cell lines that do not express Nova (Figures 8E and 8F; data not shown). pcNova-1 cotransfection with the wild-type GlyR $\alpha 2$ minigene mediated a 2.5-fold increase in E3A splicing in HeLa cells and a 2-fold increase in 293T cells. Some

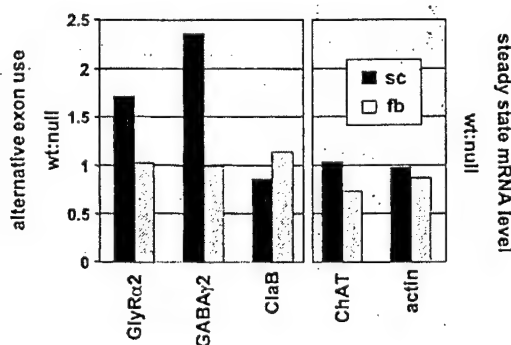


Figure 5. GlyR $\alpha 2$ and GABA $\gamma 2$ Alternative Splicing Defects Are Restricted to Nova-1-Expressing Regions of Brain in *Nova-1* Null Mice

Spinal cord or forebrain RNA was isolated from a P18 (P7 for GABA $\gamma 2$) *Nova-1* null mice and a wild-type littermate and analyzed for the indicated splicing products by RT-PCR. Data are presented as in Figure 3C. Unlike the spinal cord, quantitation of phosphorimager data using RNA from forebrain revealed no differences in the ratio of GlyR $\alpha 2$ E3A/E3B and $\gamma 2L/\gamma 2S$ GABA $\gamma 2$ exon usage between *Nova-1* null animals and wild-type littermates. ClaB alternative splicing showed no null:wild-type difference in either spinal cord or forebrain. Quantitation revealed no differences in the steady-state levels of RNA encoding the motor neuron marker ChAT or actin in *Nova-1* null spinal cord relative to wild-type.

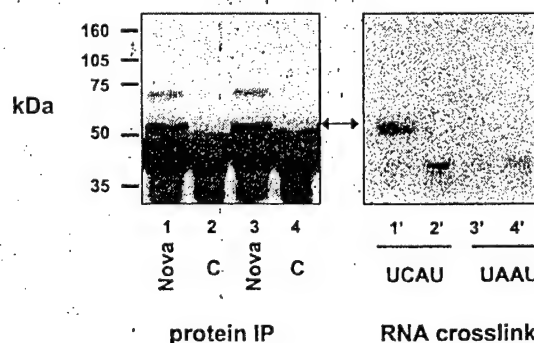


Figure 7. Native Nova Protein Binds to the GlyR $\alpha 2$ (UCAU) $_3$ Element

An extract of mouse brainstem was incubated with 90 nt ^{32}P -labeled RNAs encoding either the GlyR $\alpha 2$ UCAU intronic sequence (lanes 1, 2, 1', and 2') or mutant UAAU RNA (lanes 3, 4, 3', and 4'). Samples were UV cross-linked, RNase treated, and immunoprecipitated with either rabbit anti-Nova serum (lanes 1, 3, 1', and 3') or preimmune serum (lanes 2, 4, 2', and 4'), run on SDS-PAGE gels and transferred to a filter for Western blot analysis with POMA antiserum. The same filter was also exposed directly to X-ray film to detect ^{32}P -labeled RNA (lanes 1'-4'). The arrow indicates the molecular size of Nova-1. The strong reactivity present underneath the Nova-1 bands (lanes 1-4) is due to secondary antibody cross-reactivity with rabbit immunoglobulin G. The ~70 kDa band visible in lanes 1 and 3 is a protein related to Nova-2 that immunoprecipitates specifically with POMA antisera (Yang et al., 1998).

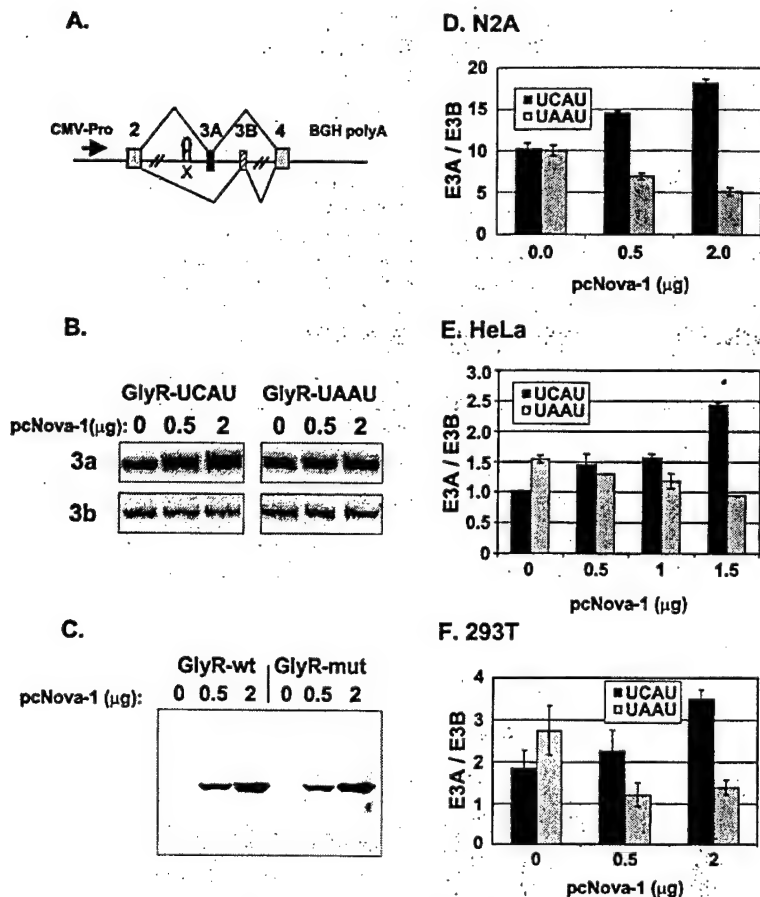


Figure 8. Nova-1 Enhances GlyR α 2 E3A Splicing in Heterologous Cell Lines

(A) The GlyR α 2 minigene used for splicing assays in transiently transfected cells contains the complete GlyR α 2 E2, a shortened intron 2, the complete genomic sequence surrounding E3A and E3B, a shortened intron 3, and 233 nucleotides at the 5' end of E4. The UCAU Nova-1 binding site in intron 2 is indicated by the schematic hairpin loop, and the UAAU mutant minigene is indicated by an X. (B) Nova-1-dependent effects on E3A splicing of the wild-type UCAU and mutant UAAU GlyR α 2 minigenes. RNA from N2A cells transfected with the wild-type or mutant GlyR α 2 minigene and the indicated amount of pcNova-1 plasmid were analyzed as in Figure 3A.

(C) Western blot using anti-T7 antibody showing the titration of T7-tagged Nova protein levels after N2A cells were cotransfected with the GlyR α 2 minigene and the indicated amounts of pcNova-1.

(D) Nova increases the ratio of E3A/E3B splicing from wild-type GlyR α 2 pre-mRNA, but this effect is abrogated in the GlyR α 2 mutant UAAU pre-mRNA. Three independent transfections were performed in which splicing products (E3A or E3B) were quantitated as in Figure 3A. Increasing amounts of pcNova-1 were cotransfected with either the wild type or mutant GlyR α 2 minigene, as indicated. Results are displayed as a ratio of E3A/E3B. The total amount of spliced product is unchanged as a result of Nova titration; the increase in E3A usage is accompanied by a compensatory decrease in E3B utilization (data not shown). Reaction products were linear with

respect to input cDNA and PCR cycle number (data not shown). Results shown are the average of three independent transfections; and error bars represent standard deviation. Similar results were obtained in two additional experiments (data not shown).

(E and F) Nova facilitates E3A splicing in HeLa (E) and 293T (F) cells specifically in wild-type UCAU GlyR α 2 transcripts. Western blot confirmed that HeLa and 293T cells did not express Nova and that Nova transfection led to linear increases in Nova protein (data not shown). Results are the average of two (E) and four (F) independent transfections, and error bars represent standard deviation.

additional effects were seen. In HeLa cells, the increase in E3A splicing was at the expense of E3B, while in 293T cells the increase in E3A-spliced product was inversely proportional to a decrease in an aberrant spliced product that contained neither E3A nor E3B (E2-E4; data not shown). We also found that in cells transfected with the mutant UAAU GlyR α 2 minigene, in the presence of Nova there was no increase in E3A splicing but again a paradoxical increase in E3B splicing, and in the absence of Nova there was a small but consistent increase in E3A splicing that is unexplained. Taken together, these experiments demonstrate that Nova-1 is able to act directly to enhance the inclusion of GlyR α 2 E3A in cell lines and that this action is dependent on the integrity of the intronic UCAU Nova binding site.

Discussion

Nova-1 Regulates Mammalian Tissue-Specific Alternative Splicing

We have combined biochemical and genetic approaches to demonstrate that Nova-1 functions to regulate alternative splicing in neurons and found that the absence of Nova-1 leads to neuronal death in spinal

and brainstem neurons. We provide strong evidence that Nova-1 functions to regulate alternative splicing in neurons by binding pre-mRNA in a sequence-specific manner to activate exon inclusion. Nova-1 present in brain extracts specifically cross-links the sequence (UCAUY)₃ but not the mutant sequence (UAAUY)₃, consistent with previous coimmunoprecipitation experiments demonstrating that Nova-1 protein binds GlyR α 2 pre-mRNA in vivo. In cotransfection assays, GlyR α 2 pre-mRNAs harboring this (UCAUY)₃ sequence show a Nova-1 dose-dependent increase in utilization of the alternatively spliced E3A, while mutant pre-mRNAs harboring the (UAAUY)₃ point mutant do not show this effect. In Nova-1 null mice, there is a reciprocal 2-fold decrease in the ratio of GlyR α 2 E3A relative to E3B splicing. These data provide converging lines of evidence that Nova-1 regulates GlyR α 2 splicing by directly binding to the pre-mRNA in neurons.

We also find a second specific defect in alternative splicing in Nova-1 null mice, which is deficient inclusion of the GABA α γ 2L exon. The defective exon utilization of γ 2L and GlyR α 2 E3A in Nova-1 null mice, together with the finding that Nova-1 enhanced E3A utilization in heterologous cells, suggests that Nova-1 most likely

acts to activate GlyR α 2 E3A and GABA γ 2L exon inclusion in neurons. An alternative, and not entirely exclusive, possibility is that Nova-1 activates splicing indirectly by competing with a splicing inhibitor. Such an interpretation would be consistent with the observation that Nova action can be antagonized by a newly discovered brain-enriched variant of PTB (A. D. Polydorides et al., submitted). However, a strictly indirect action of Nova appears unlikely, since the action of Nova-1 in transfected cells is abolished by mutation in an intronic RNA element.

Several unanswered questions regarding tissue-specific splicing are addressed by the finding that Nova-1 regulates alternative splicing in neurons. RNA-binding proteins such as Sxl and Tra/Tra-2 in *Drosophila*, identified by genetic approaches, act in a sequence-specific manner to regulate alternative splicing of specific target transcripts. However, genetics has not previously been applied to the study of RNA-binding proteins in mammals, and biochemical studies applied to complex splicing problems have led not to the identification of tissue-specific splicing factors, but to the identification of ubiquitously expressed RNA-binding proteins. These studies have yielded one clear model in which differential regulation of splicing in mammals is generated through relatively subtle changes in the levels of such ubiquitous splicing factors. For example, changing the ratio of the constitutively expressed hnRNP protein hnRNP A1 to the SR protein ASF-SF2 mediates quantitative changes, ranging from 1.25-fold to severalfold in cell transfection studies, in the utilization of proximal versus distal splice site choices at alternatively spliced exons (Caceres et al., 1994). Our data demonstrate that in addition to the role general factors play in regulating alternative splicing, tissue-specific factors that recognize specific RNA ligands exist in mammals. A working model for Nova action is that it acts to enhance utilization of adjacent exons by recruiting general factors, such as SR proteins, to promote the assembly of a splicing-competent complex.

Nova-1 Regulates Neuron-Specific Alternative Splicing

Three families of strictly neuron-specific RNA-binding proteins have been identified. Musashi proteins are expressed in mitotic neuronal precursors (Sakakibara et al., 1996) and play a role in the early development of the nervous system in *Drosophila*. Nova (Buckanovich et al., 1993; Yang et al., 1998) and Hu (Szabo et al., 1991) proteins were identified as PND target antigens. These autoimmune neurologic disorders are believed to be triggered when proteins normally sequestered from immune surveillance in neurons are ectopically expressed in tumor cells (Darnell, 1996, 1998), leading to the development of antigen-specific antibodies and T cells (Darnell, 1996; Albert et al., 1998, 1999).

Detailed analysis of the developmental and adult expression patterns of both Nova (Buckanovich et al., 1993, 1996) and Hu (Okano and Darnell, 1997; Wakamatsu and Weston, 1997) antigens revealed that their expression is entirely restricted to neurons throughout development. Nova-1 expression is not only specific to neurons, but to a subset of neurons in the hypothalamus,

brainstem, and spinal cord (Buckanovich et al., 1993, 1996; Figures 1 and 2). Therefore, the identification of Nova-1 as a splicing factor provides a mechanism by which individual neurons can differentially regulate splicing in a manner qualitatively distinct from that of other cell types.

In previous studies of the mechanism of neuron-specific splicing, intronic enhancer elements similar in nature to the GlyR α 2 (UCAUY)₃ sequences have been identified. These include multiple elements that promote neuron-specific exon inclusion in the c-src, calcitonin/CGRP, GABA γ 2, and agrin genes (Black, 1992; Zhang et al., 1996; Wei et al., 1997; Lou et al., 1999). Regulation of splicing appears complex given that multiple elements, both positive and negative, influence alternative splicing events. For example, in c-src, in addition to positive acting signals identified as targets of an RNA-binding protein complex involving KSRP, hnRNP F, and hnRNP H (Min et al., 1997; Chou et al., 1999), negative elements are found in the polypyrimidine tract, upstream of the neuronal n-src exon, that interact with PTB (Chan and Black, 1997) and downstream of the n-src exon that interact with brain-enriched PTB (D. L. Black et al., submitted).

The mechanism of neuron-specific splicing in the GABA γ 2 gene has similar complexities (Ashiya et al., 1995; Zhang et al., 1996, 1999). A number of negative elements surrounding the γ 2 exon were found to bind PTB, and evidence was also found for positive acting splicing elements. These observations, together with our data, suggest that Nova-1 may act in concert with general factors such as PTB or brain-enriched PTB to regulate GABA γ 2L exon utilization, which would be consistent with the finding that Nova interacts with brain PTB (A. D. Polydorides et al., personal communication). GABA γ 2L is preferentially included in the hypothalamus, midbrain, brainstem, and spinal cord (Zhang et al., 1996), and this correlates precisely with the pattern of Nova-1 expression, supporting our finding of a role for Nova-1 in generating the γ 2L splice variant in vivo.

Nova-1, Neuronal Death, and Neurologic Disease

Several recent reports have implicated splicing defects as a cause of motor neuron death in the spinal cord. Dreyfuss and colleagues have determined that a protein mutated in the human disorder spinal muscular atrophy plays a critical role in the generation of snRNPs and have suggested that the proper regulation of splicing is necessary for postnatal survival of motor neurons (Fischer et al., 1997). In addition, Lin et al. (1998) have reported that splicing defects in the glutamate transporter EAAT2 are associated with the motor neuron disorder amyotrophic lateral sclerosis. The phenotype of Nova-1 null mice, which phenocopies many aspects of the human POMA syndrome (see below), together with the biochemical evidence presented here, provides a particularly compelling case illustrating splicing defects that are associated with neuronal death in the ventral spinal cord.

Our results demonstrate that Nova-1 is essential for neuronal viability in mature neurons in postnatal mice.

Nova-1 null neurons appear to undergo normal morphogenesis, assessed histochemically, and axon pathfinding, assessed by electron microscopy of the neuromuscular junction at birth (J. Sanes and R. B. D., unpublished data). However, the TUNEL-positive and pyknotic nuclei we observed in P5 *Nova-1* null neurons reflect the end stage of apoptotic neuronal death, and it is difficult to be certain at what age the neurons become dysfunctional. Moreover, we cannot rule out the possibility that *Nova-1* may also play a role in early neuronal development. For example, *Nova-2* is expressed together with *Nova-1* in early postmitotic neurons and is downregulated in *Nova-1*-expressing neurons postnatally, suggesting that *Nova-2* could rescue *Nova-1* function before birth. The defect in neuronal viability suggests either a loss of function, such as a failure of mechanisms necessary to ensure survival of mature neurons, or a gain of function, such as acquisition of a new toxic activity that mediates neuronal death in *Nova-1* null neurons. For example, aberrant splicing of the α_{1A} calcium channel in mice is believed to result in a gain-of-function mutation, which leads to a delayed onset neurologic disorder manifest by ataxia, motor seizures, and absence seizures (Fletcher et al., 1996). In *Nova-1* null mice, defects in inhibitory receptor splicing could lead to a gain of function, such that glycine and/or GABA receptors become toxic, for example, by creating imbalances between the neuronal inhibition and excitation that lead to excitotoxic cell death.

Patients with POMA have symptoms that suggest abnormal inhibitory control over brainstem and spinal motor pathways (Luque et al., 1991). For example, opsoclonus is thought to result from failure of normal functioning of brainstem inhibitory glycinergic neurons, termed omipause neurons (Averbuch-Heller and Remler, 1996), and brainstem or spinal myoclonus is believed to relate to a failure of inhibition of motor neurons and may involve GABAergic or glycinergic systems (Rothwell, 1995; Caviness, 1996; Rajendra et al., 1997). Moreover, mice with the spasmodic mutation (Ryan et al., 1994; Saul et al., 1994) and humans with hyperekplexia (Shiang et al., 1993) harbor defects in inhibitory glycine receptors and suffer from myoclonic syndromes that bear marked similarities to the motor symptoms evident in *Nova-1* null mice and in POMA patients. Thus, the defects in splicing of the inhibitory glycine and GABA receptors in *Nova-1* null mice suggest a reasonable model for the mechanism of motor dysfunction in both the murine and human systems. In POMA, immunologic targeting of *Nova* protein may lead to defects in inhibitory receptor function and thereby to the excess motor activity evident in the disorder.

Complexity of Regulation of Neuronal Alternative Splicing

The identification of the glycine and GABA_A receptor pre-mRNAs as *Nova-1* targets raises the question of the extent to which *Nova* proteins are used to regulate alternative splicing in neurons. The *Nova* gene family itself consists of at least two (Buckanovich et al., 1993; Yang et al., 1998) and possibly additional *Nova* gene family members (Fletcher et al., 1997; Yang et al., 1998; see also Figure 1), and the *Nova-1* and *Nova-2* transcripts are themselves alternatively spliced (Buckanovich et al., 1993; Yang et al., 1998). Evaluation of *Nova*

protein-protein interactions using biochemical and yeast two-hybrid assays suggests that *Nova-1* and *Nova-2* can form homo- and heterotypic protein interactions (Lewis et al., 1999; R. B. D. et al., unpublished data). Moreover, data from the crystal structure of *Nova* KH3 (Lewis et al., 1999) and the *Nova* KH3-RNA complex (Lewis et al., 2000) demonstrate that *Nova* proteins dimerize, even when bound to RNA.

Neurons are likely to utilize some immune system-like mechanism to generate diversity from a relatively small number of primary transcripts (Darnell, 1998). The significance of a neuron-specific system that regulates alternative splicing is that it provides a means for the generation of such diversity. While genome rearrangements might occur in the brain (see Gao et al., 1998; Kohmura et al., 1998; Serafini, 1999; Wu and Maniatis, 1999), to date alternative splicing and RNA editing are the only relevant mechanisms known to operate in neurons. *Nova* proteins, acting in concert with other proteins, are likely to play an important role in generating the complexity of neuronal function.

Experimental Procedures

Plasmid Constructs

The GlyR α 2 minigene (cloned in pCDNA3, Invitrogen) corresponds to the following GlyR α 2 E2 mouse genomic sequences: bases 81–276 (GenBank accession number X75842), 63–583 (GenBank accession number X75843), and 10–300 (GenBank accession number 75844). The mutated GlyR α 2 minigene, in which the *Nova-1* binding site 5'-UCAUCAUUAUUUCAUUUUUUUU-3' was mutated to 5'-UAAUAAUUUUUAAUUAAGGUU-3', was generated by PCR. A cDNA encoding T7-tagged *Nova-1* was cloned into pcXHook (Invitrogen). All DNA constructs generated by PCR were sequenced in their entirety.

PCR Primers

The primers used were E2 F, 5'-AGCTTTCTGCAAGACCATGAC; E4 R, 5'-GAAGATCTCCAAATCCAAGGAATCATCTGGG; α 2Gly R E3A, 5'-CATGGTGGTTTCTGTGACTGATC; α 2Gly R E3B, 5'-CATTG TAGTTTCTGCTATTGACCCAAAG; α 2Gly R E4, 5'-TCCAAATCCAA GGAATCATCTGGG; actin F, 5'-GTGGCCGCTCTAGGCACCA; actin R, CCCCCCTGAACCCCTAAGGCCAACCG; mGluR1 F, 5'-CCT GGGGTGCATGTTTACTCC; mGluR1 R, 5'-AGGCCGTCTCGTTGGT CTTC; ICH-1 F, 5'-GTCTCATCTTCATCAACTCC; ICH-1 R, 5'-ATG CTAACCTGCTCAAGTCTA; ClaB F, 5'-ACCGAACAGGAGTGGCG GGAG; ClaB R, 5'-GGGGTCTCCTCTGGATTCT; γ 2 GABA F, 5'-GTATGGCACCCTGCATTATTTTGTG; γ 2 GABA R, 5'-TTGAATG GTTGCTGATCTGGGACG; ChAT F, 5'-ATGCCTATCCTGGAAAA GTTCCG; ChAT R, 5'-AGTGCTCCGAGCAAAGATCACAG; src F, 5'-CCAAGCTCTTCGGAGGCTTCAACTC; src R, 5'-CACATAGTTGC TGGGGATGTAACCG; agrin F, 5'-GGGATAGTTGAGAAGTCACTGG GGG; and agrin R, 5'-CGAAGCCAGCGTTGGTGTG.

Generation of Targeted Mutation

Genomic clones encoding *Nova-1* were isolated from a phage library derived from a 129 mouse (Stratagene). A 1.8 kb XbaI-AlwNI fragment encoding sequences upstream of the initiator methionine and a 6.5 kb SacI fragment harboring sequences downstream of the second coding exon were cloned into pBS⁺ (Stratagene). A PacI linker was ligated to a unique NotI site between these fragments to allow insert of the IRES-*lacZ* LTNL cassette (Mombaerts et al., 1996). Plasmid was linearized with SalI and used to electroporate ES cells, which were grown as described (Mombaerts et al., 1996), including 150 μ g/ml of G418 starting 24 hr after transfection. Genomic DNA from individual colonies was digested with EcoRI and screened by Southern blot with a 5' *Nova-1* genomic probe (0.5 kb SacI-XbaI fragment).

Clones that had undergone homologous recombination were transfected with a plasmid expressing Cre recombinase (pB5185) to excise the LTNL cassette. Accurate homologous recombination in positive clones was confirmed by Southern blot with HindIII-digested genomic DNA and a 3' Nova-1 genomic probe (0.75 kb SacI fragment). A homologous recombinant clone was injected into C57/BL6 blastocysts to produce germline chimeras.

Histology and TUNEL Assay

The mice were perfused with 4% buffered formaldehyde, and the brains were removed, postfixed overnight, and embedded in paraffin blocks. Tissue sections (14 μ m) were deparaffinized, rehydrated, stained with hematoxylin and eosin, and visualized by light microscopy using a Zeiss Axioplan microscope. Pyknotic neurons were quantitated by counting all pyknotic cells in each of 27 14 μ m thick sections of spinal cord from a single Nova-1 null mouse or a single wild-type littermate. For X-gal staining, whole-mount staining was performed as described (Mombaerts et al., 1996).

Apoptosis was assessed by the DNA terminal transferase nick end translation method, or TUNEL assay, as described previously (Fuks et al., 1995), and visualized by fluorescence microscopy.

RNA Preparation and RT-PCR

Purified RNA (Chomczynski and Sacchi, 1987) was reverse transcribed using random hexamers, and cDNA products were amplified using PfuTurbo (Stratagene) with 40 pmol of each primer and 0.5 pmol of one γ -³²P-ATP-labeled primer. PCR products were confirmed by sequencing. PCR was linear with respect to input cDNA and cycle number (data not shown).

T7 Transcription and UV Cross-Linking

RNAs for crosslinking were generated by PCR using the wild-type or mutant GlyR α 2 minigene, followed by *in vitro* transcription. The primers used were T7primer1 (sense) (5'-AGTAATACGACTCACTATAGGGATCATGCAGTCTCTGGTTAAT) and primer2 (antisense) (5'-AGGTCCATCAACATCTGTGG), which amplified a 90 nt fragment surrounding the Nova-1 binding site. Gel-purified PCR products were used as templates for *in vitro* transcription. Thirty microliters of total cell extracts from mouse brainstem (50 μ l/ μ g) were irradiated with UV light for 15 min. The samples were treated with 80 U of RNase A (Worthington) for 30 min at 37°C. UV cross-linking reactions were immunoprecipitated with a 5 μ l polyclonal rabbit anti-Nova-1 antisera in a total volume of 100 μ l of lysis buffer. The immune complexes were precipitated with protein A-Sepharose beads (Sigma), run on 10% SDS-PAGE gels, transferred to nitrocellulose, and exposed to film.

Cell Transfection

N2A, HeLa, and 293T cells were grown to 60% confluence in 35 mm dishes or in 6-well plates. 2.5 μ g of total DNA, comprised of 0.25 μ g of the wild-type or mutant pcDNA3 GlyR α 2 minigene, 0.25 μ g of pCMV β -galactosidase, and variable amounts of pcNova-1 and empty pcXHook vector, were incubated with 95 μ l of Dulbecco's modified Eagle's medium (DMEM) and 5 μ l of Eugene (Roche) for 15 min at room temperature. The mixture was then added to the cells in a total volume of 2 ml 10% FBS-DMEM. After 48 hr, the cells were collected and used for RNA extraction, protein extraction, and β -galactosidase assay. Western blot analysis and β -galactosidase assays were performed as described (Okano et al., 1999).

Acknowledgments

The authors are grateful to Peter Mombaerts for advice on generating Nova-1 null mice and for the IRES- τ -lacZ LTNL cassette and to Nathaniel Heintz for critical review of the manuscript. This work was supported by the National Institutes of Health (RO1 NS34389), an Irma T. Hirschl Career Scientist Award (R. B. D.), the Breast Cancer Research Program (DAMD17-97-1-7097) (K. B. J.), and the Ataxia Telangiectasia Children's Project (H. J. O.).

Received August 16, 1999; revised January 7, 2000.

References

Albert, M.L., Damell, J.C., Bender, A., Francisco, L.M., Bhardwaj, N., and Damell, R.B. (1998). Tumor-specific killer cells in paraneoplastic cerebellar degeneration. *Nat. Med.* 4, 1321-1324.

Albert, M.L., Austin, L.M., and Damell, R.B. (1999). Detection and treatment of activated T cells in the cerebrospinal fluid of patients with paraneoplastic cerebellar degeneration. *Ann. Neurol.*, in press.

Amara, S.G., Jonas, V., Rosenfeld, M.G., Ong, E.S., and Evans, R. (1982). Alternative RNA processing in calcitonin gene expression generates mRNAs encoding different polypeptide products. *Nature* 298, 240-244.

Aming, S., Grüter, P., Bilbe, G., and Kramer, A. (1996). Mammalian splicing factor SF1 is encoded by variant cDNAs and binds to RNA. *RNA* 2, 794-810.

Ashiya, M., Zhang, L., and Grabowski, P.J. (1995). Regulated splicing of gamma α 2 pre-messenger RNA in neuronal cells. *Nucleic Acids Symp. Ser.* 33, 215-216.

Ashley, C.T., Wilkinson, K.D., Reines, D., and Warren, S.T. (1993). FMR-1 protein: conserved RNP family domains and selective RNA binding. *Science* 262, 563-566.

Averbuch-Heller, L., and Remler, B. (1996). Opsoclonus. *Semin. Neurol.* 16, 21-26.

Baker, B.S. (1989). Sex in flies: the splice of life. *Nature* 340, 521-524.

Bechade, C., Sur, C., and Triller, A. (1994). The inhibitory neuronal glycine receptor. *Bioessays* 16, 735-744.

Black, D.L. (1992). Activation of c-src neuron-specific splicing by an unusual RNA element *in vivo* and *in vitro*. *Cell* 69, 795-807.

Black, D.L. (1998). Splicing in the inner ear: a familiar tune, but what are the instruments? *Neuron* 20, 165-168.

Buckanovich, R.J., and Damell, R.B. (1997). The neuronal RNA binding protein Nova-1 recognizes specific RNA targets *in vitro* and *in vivo*. *Mol. Cell. Biol.* 17, 3194-3201.

Buckanovich, R.J., Posner, J.B., and Damell, R.B. (1993). Nova, the paraneoplastic RI antigen, is homologous to an RNA-binding protein and is specifically expressed in the developing motor system. *Neuron* 11, 657-672.

Buckanovich, R.J., Yang, Y.Y., and Damell, R.B. (1996). The onconeural antigen Nova-1 is a neuron-specific RNA-binding protein, the activity of which is inhibited by paraneoplastic antibodies. *J. Neurosci.* 16, 1114-1122.

Burd, C.G., and Dreyfuss, G. (1994). Conserved structures and diversity of functions of RNA-binding proteins. *Science* 265, 615-621.

Caceres, J.F., Stamm, S., Helfan, D.M., and Krainer, A.R. (1994). Regulation of alternative splicing *in vivo* by overexpression of antagonistic splicing factors. *Science* 265, 1706-1709.

Caviness, J.N. (1996). Myoclonus. *Mayo Clin. Proc.* 71, 679-688.

Chan, R.C., and Black, D.L. (1997). The polypyrimidine tract binding protein binds upstream of neural cell-specific c-src exon N1 to repress the splicing of the intron downstream. *Mol. Cell. Biol.* 17, 4667-4676.

Chomczynski, P., and Sacchi, N. (1987). Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. *Anal. Biochem.* 162, 156-159.

Chou, M.Y., Rooke, N., Turck, C.W., and Black, D.L. (1999). hnRNP H is a component of a splicing enhancer complex that activates a c-src alternative exon in neuronal cells. *Mol. Cell. Biol.* 19, 69-77.

Claeysen, S., Faye, P., Sebben, M., Taviaux, S., Bockaert, J., and Dumuis, A. (1998). 5-HT $_4$ receptors: cloning and expression of new splice variants. *Ann. NY Acad. Sci.* 861, 49-56.

Conn, P.J., and Pin, J.P. (1997). Pharmacology and functions of metabotropic glutamate receptors. *Annu. Rev. Pharmacol. Toxicol.* 37, 205-237.

Damell, R.B. (1996). Onconeural antigens and the paraneoplastic neurologic disorders: at the intersection of cancer, immunity and the brain. *Proc. Natl. Acad. Sci. USA* 93, 4529-4536.

Damell, R.B. (1998). Immunologic complexity in neurons. *Neuron* 21, 947-950.

Ehlers, M.D., Tingley, W.G., and Haganir, R.L. (1995). Regulated subcellular distribution of the NR1 subunit of the NMDA receptor. *Science* 269, 1734-1737.

Ehlers, M.D., Fung, E.T., O'Brien, R.J., and Haganir, R.L. (1998). Splice variant-specific interaction of the NMDA receptor subunit NR1 with neuronal intermediate filaments. *J. Neurosci.* 18, 720-730.

Engelbrecht, J., and Roeder, G.S. (1990). MER1, a yeast gene required for chromosome pairing and genetic recombination, is induced in meiosis. *Mol. Cell. Biol.* 10, 2379-2389.

Ferns, M., Hoch, W., Campanelli, J.T., Rupp, F., Hall, Z.W., and Scheller, R.H. (1992). RNA splicing regulates agrin-mediated acetylcholine receptor clustering activity on cultured myotubes. *Neuron* 8, 1079-1086.

Fischer, U., Liu, Q., and Dreyfuss, G. (1997). The SMN-SIP1 complex has an essential role in spliceosomal snRNP biogenesis. *Cell* 90, 1023-1029.

Fletcher, C.F., Lutz, C.M., O'Sullivan, T.N., Shaughnessy, J.D., Jr., Hawkes, R., Frankel, W.N., Copeland, N.G., and Jenkins, N.A. (1996). Absence epilepsy in tottering mutant mice is associated with calcium channel defects. *Cell* 87, 607-617.

Fletcher, C.F., Okano, H.J., Gilbert, D.J., Yang, Y.Y.L., Yang, C.W., Copeland, N.G., Jenkins, N.A., and Darnell, R.B. (1997). Murine chromosomal locations of nine genes encoding homologs of human paraneoplastic neurologic disorder antigens. *Genomics*, in press.

Foley, K.P., Leonard, M.W., and Engel, J.D. (1993). Quantitation of RNA using the polymerase chain reaction. *Trends Genet.* 9, 380-385.

Fuks, Z., Alfieri, A., Haimovitz-Friedman, A., Seddon, A., and Cordon-Cardo, C. (1995). Intravenous basic fibroblast growth factor protects the lung but not mediastinal organs against radiation-induced apoptosis in vivo. *Cancer J. Sci. Am.* 1, 62.

Gao, Y., Sun, Y., Frank, K.M., Dikkes, P., Fujiwara, Y., Seidl, K.J., Sekiguchi, J.M., Rathbun, G.A., Swat, W., Wang, J. et al. (1998). A critical role for DNA end-joining proteins in both lymphogenesis and neurogenesis. *Cell* 95, 891-902.

Grabowski, P.J. (1998). Splicing regulation in neurons: tinkering with cell-specific control. *Cell* 92, 709-712.

Hertel, K.J., Lynch, K.W., and Maniatis, T. (1997). Common themes in the function of transcription and splicing enhancers. *Curr. Opin. Cell Biol.* 9, 350-357.

Hollmann, M., and Heinemann, S. (1994). Cloned glutamate receptors. *Annu. Rev. Neurosci.* 17, 31-108.

Hollmann, M., Boulter, J., Maron, C., Beasley, L., Sullivan, J., Pecht, G., and Heinemann, S. (1993). Zinc potentiates agonist-induced currents at certain splice variants of the NMDA receptor. *Neuron* 10, 943-954.

Horowitz, D.S., and Krainer, A.R. (1994). Mechanisms for selecting 5' splice sites in mammalian pre-mRNA splicing. *Trends Genet.* 10, 100-106.

Kohmura, N., Senzaki, K., Hamada, S., Kai, N., Yasuda, R., Watanabe, M., Ishii, H., Yasuda, M., Mishina, M., and Yagi, T. (1998). Diversity revealed by a novel family of cadherins expressed in neurons at a synaptic complex. *Neuron* 20, 1137-1151.

Kuhse, J., Kuryatov, A., Maulet, Y., Malosio, M.L., Schmieden, V., and Betz, H. (1991). Alternative splicing generates two isoforms of the $\alpha 2$ subunit of the inhibitory glycine receptor. *FEBS Lett.* 283, 73-77.

Lewis, H., Musunuru, K., Jensen, K.B., Carne, E., Chen, H., Darnell, R.B., and Burley, S.K. (2000). Sequence-specific RNA binding by a Nova KH-domain: implications for paraneoplastic disease and the fragile-X syndrome. *Cell*, in press.

Lewis, H.A., Chen, H., Edo, C., Buckanovich, R.J., Yang, Y.Y., Musunuru, K., Zhong, R., Darnell, R.B., and Burley, S.K. (1999). Crystal structures of Nova-1 and Nova-2 K-homology RNA-binding domains. *Structure* 7, 191-203.

Lin, C.L.G., Bristol, L.A., Jin, L., Dykes-Hoberg, M., Crawford, T., Clawson, L., and Rothstein, J.D. (1998). Aberrant RNA processing in a neurodegenerative disease: the cause for absent EAAT2, a glutamate transporter, in amyotrophic lateral sclerosis. *Neuron* 20, 589-602.

Lou, H., Helfman, D.M., Gagel, R.F., and Berget, S.M. (1999). Polypyrimidine tract-binding protein positively regulates inclusion of an alternative 3'-terminal exon. *Mol. Cell. Biol.* 19, 78-85.

Luque, F., Fumeaux, H., Ferziger, R., Rosenblum, M., Wray, S., Schold, S., Glantz, M., Jaekle, K., Biran, H., Lesser, M., et al. (1991). Anti-Ri: an antibody associated with paraneoplastic opsoclonus and breast cancer. *Ann. Neurol.* 29, 241-251.

Macdonald, R.L. (1995). Ethanol, gamma-aminobutyrate type A receptors, and protein kinase C phosphorylation. *Proc. Natl. Acad. Sci. USA* 92, 3633-3635.

Martinez, R., Mathey-Prevot, B., Bernards, A., and Baltimore, D. (1987). Neuronal pp60c-src contains a six-amino acid insertion relative to its non-neuronal counterpart. *Science* 237, 411-415.

Michael, W.M., Eder, P.S., and Dreyfuss, G. (1997). The K nuclear shuttling domain: a novel signal for nuclear import and nuclear export in the hnRNP K protein. *EMBO J.* 16, 3587-3598.

Min, H., Turck, C.W., Nikolic, J.M., and Black, D.L. (1997). A new regulatory protein, KSRP, mediates exon inclusion through an intronic splicing enhancer. *Genes Dev.* 11, 1023-1036.

Missler, M., and Südhof, T.C. (1998). Neurexins: three genes and 1001 products. *Trends Genet.* 14, 20-26.

Mombaerts, P., Wang, F., Dulac, C., Chao, S.K., Nemes, A., Mendelsohn, M., Edmondson, J., and Axel, R. (1996). Visualizing an olfactory sensory map. *Cell* 87, 675-686.

Okano, H.J., and Darnell, R.B. (1997). A hierarchy of Hu RNA binding proteins in developing and adult neurons. *J. Neurosci.* 17, 3024-3037.

Okano, H.J., Park, W.Y., Corradi, J.P., and Darnell, R.B. (1999). The cytoplasmic Purkinje antigen cdr2 downregulates Myc function: implications for neuronal and tumor cell survival. *Genes Dev.* 13, 2087-2098.

Ostareck, D.H., Ostareck-Lederer, A., Wilm, M., Thiele, B.J., Mann, M., and Hentze, M.W. (1997). mRNA silencing in erythroid differentiation—hnRNP K and hnRNP E1 regulate 15-lipoxygenase translation from the 3' end. *Cell* 89, 597-606.

Ostareck-Lederer, A., Ostareck, D.H., and Hentze, M.W. (1998). Cytoplasmic regulatory functions of the KH-domain proteins hnRNPs K and E1/E2. *Trends Biochem. Sci.* 23, 409-411.

Petersen-Mahrt, S.K., Estmer, C., Ohmalm, C., Matthews, D.A., Russell, W.C., and Akusjarvi, G. (1999). The splicing factor-associated protein, p32, regulates RNA splicing by inhibiting ASF/SF2 RNA binding and phosphorylation. *EMBO J.* 18, 1014-1024.

Picetti, R., Saiardi, A., Abdel Samad, T., Bozzi, Y., Baik, J.H., and Borrelli, E. (1997). Dopamine D2 receptors in signal transduction and behavior. *Crit. Rev. Neurobiol.* 11, 121-142.

Rajendra, S., Lynch, J.W., and Schofield, P.R. (1997). The glycine receptor. *Pharmacol. Ther.* 73, 121-46.

Rothwell, J.C. (1995). Brainstem myoclonus. *Clin. Neurosci.* 3, 214-218.

Ryan, S.G., Buckwalter, M.S., Lynch, J.W., Handford, C.A., Segura, L., Shiang, R., Wasmuth, J.J., Camper, S.A., Schofield, P., and O'Connell, P. (1994). A missense mutation in the gene encoding the alpha 1 subunit of the inhibitory glycine receptor in the spasmodic mouse. *Nat. Genet.* 7, 131-135.

Sakakibara, S., Imai, T., Hamaguchi, K., Okabe, M., Aruga, J., Nakajima, K., Yasutomi, D., Nagata, T., Kurihara, Y., Uesugi, S., et al. (1996). Mouse-musashi-1, a neural RNA binding protein highly enriched in the mammalian CNS stem cell. *Dev. Biol.* 176, 230-242.

Saul, B., Schmieden, V., Kling, C., Mulhardt, C., Gass, P., Kuhse, J., and Becker, C.M. (1994). Point mutation of glycine receptor alpha 1 subunit in the spasmodic mouse affects agonist responses. *FEBS Lett.* 350, 71-76.

Serafini, T. (1999). Finding a partner in a crowd: neuronal diversity and synaptogenesis. *Cell* 98, 133-136.

Shiang, R., Ryan, S.G., Zhu, Y.Z., Hahn, A.F., O'Connell, P., and Wasmuth, J.J. (1993). Mutations in the alpha 1 subunit of the inhibitory glycine receptor cause the dominant neurologic disorder, hyperekplexia. *Nat. Genet.* 5, 351-358.

Siebel, C.W., Admon, A., and Rio, D.C. (1995). Soma-specific expression and cloning of PSI, a negative regulator of P element pre-mRNA splicing. *Genes Dev.* 9, 269-283.

Singer, J.H., Talley, E.M., Bayliss, D.A., and Berger, A.J. (1998). Development of glycinergic synaptic transmission to rat brain stem motoneurons. *J. Neurophysiol.* 80, 2608-2620.

Siomi, H., Siomi, M., Nussbaum, R., and Dreyfuss, G. (1993). The protein product of the fragile X gene, FMR1, has characteristics of an RNA-binding protein. *Cell* 74, 291-298.

ED: PG RANGE?

- Sommer, B., Keinänen, K., Verdoorn, T.A., Wisden, W., Burnashev, N., Herb, A., Kohler, M., Takagi, T., Sakman, B., and Seeburg, P.H. (1990). Flip and flop: a cell-specific functional switch in glutamate-operated channels of the CNS. *Science* 249, 1580-1585.
- Szabo, A., Dalmau, J., Manley, G., Rosenfeld, M., Wong, E., Henson, J., Posner, J.B., and Fumeaux, H.M. (1991). HuD, a paraneoplastic encephalomyelitis antigen contains RNA-binding domains and is homologous to Elav and sex lethal. *Cell* 67, 325-333.
- Tacke, R., and Manley, J.L. (1999). Functions of SR and Tra2 proteins in pre-mRNA splicing regulation. *Proc. Soc. Exp. Biol. Med.* 220, 59-63.
- Tian, M., and Maniatis, T. (1992). Positive control of pre-mRNA splicing in vitro. *Science* 256, 237-240.
- Ullrich, B., Ushkaryov, Y.A., and Südhof, T.C. (1995). Cartography of neuroligins: more than 1000 isoforms generated by alternative splicing and expressed in distinct subsets of neurons. *Neuron* 14, 497-507.
- Valcarcel, J., Singh, R., Zamore, P.D., and Green, M.R. (1993). The protein Sex-lethal antagonizes the splicing factor U2AF to regulate alternative splicing of transformer pre-mRNA. *Nature* 362, 171-175.
- Wakamatsu, Y., and Weston, J.A. (1997). Sequential expression and role of Hu RNA-binding proteins during neurogenesis. *Development* 124, 3449-3460.
- Wang, L., Miura, M., Bergeron, L., Zhu, H., and Yuan, J. (1994). Ich-1, an I κ B/ced-3-related gene, encodes both positive and negative regulators of programmed cell death. *Cell* 78, 739-750.
- Wang, Z., and Grabowski, P.J. (1996). Cell- and stage-specific splicing events resolved in specialized neurons of the rat cerebellum. *RNA* 2, 1241-1253.
- Wei, N., Lin, C.Q., Modafferi, E.F., Gomes, W.A., and Black, D.L. (1997). A unique intronic splicing enhancer controls the inclusion of the agrin Y exon. *RNA* 3, 1275-1288.
- Wu, Q., and Maniatis, T. (1999). A striking organization of a large family of human neural cadherin-like cell adhesion genes. *Cell* 97, 779-790.
- Yang, Y.Y.L., Yin, G.L., and Darnell, R.B. (1998). The neuronal RNA binding protein Nova-2 is implicated as the autoantigen targeted in POMA patients with dementia. *Proc. Natl. Acad. Sci. USA* 95, 13254-13259.
- Zaki, P.A., Bilsky, E.J., Vanderah, T.W., Lai, J., Evans, C.J., and Porreca, F. (1996). Opioid receptor types and subtypes: the delta receptor as a model. *Annu. Rev. Pharmacol. Toxicol.* 36, 379-401.
- Zhang, L., Ashiya, M., Sherman, T.G., and Grabowski, P.J. (1996). Essential nucleotides direct neuron-specific splicing of gamma 2 pre-mRNA. *RNA* 2, 682-698.
- Zhang, L., Liu, W., and Grabowski, P.J. (1999). Coordinate repression of a trio of neuron-specific splicing events by the splicing regulator PTB. *RNA* 5, 117-130.

Sequence-Specific RNA Binding by a Nova KH Domain: Implications for Paraneoplastic Disease and the Fragile X Syndrome

Hal A. Lewis,*^{||} Kiran Musunuru,*[†]
Kirk B. Jensen,[†] Carme Edo,* Hua Chen,*[‡]
Robert B. Darnell,[‡] and Stephen K. Burley*^{‡§}

*Laboratories of Molecular Biophysics

[†]Laboratory of Molecular Neuro-Oncology

[‡]Howard Hughes Medical Institute

The Rockefeller University

1230 York Avenue

New York, New York 10021

Summary

The structure of a Nova protein K homology (KH) domain recognizing single-stranded RNA has been determined at 2.4 Å resolution. Mammalian Nova antigens (1 and 2) constitute an important family of regulators of RNA metabolism in neurons, first identified using sera from cancer patients with the autoimmune disorder paraneoplastic opsoclonus-myoclonus ataxia (POMA). The structure of the third KH domain (KH3) of Nova-2 bound to a stem loop RNA resembles a molecular vise, with 5'-Ura-Cyt-Ade-Cyt-3' pinioned between an invariant Gly-X-X-Gly motif and the variable loop. Tetranucleotide recognition is supported by an aliphatic α helix/ β sheet RNA-binding platform, which mimics 5'-Ura-Gua-3' by making Watson-Crick-like hydrogen bonds with 5'-Cyt-Ade-3'. Sequence conservation suggests that fragile X mental retardation results from perturbation of RNA binding by the FMR1 protein.

Introduction

Protein-RNA interactions are central to a host of important cellular processes, including protein biosynthesis, pre-mRNA processing, RNA localization, and mRNA stabilization and degradation (Nagai, 1996). Several conserved protein sequence motifs implicated in RNA binding have been identified over the past decade (Burd and Dreyfuss, 1994), one of the most common being the K homology (KH) motif, the subject of this paper. Originally identified as a conserved region represented three times in the hnRNP K protein (Siomi et al., 1993a), KH motifs span about 70 residues with a characteristic pattern of hydrophobic residues, an invariant Gly-X-X-Gly segment, and a variable loop. Since their initial detection, a superfamily of more than 100 KH proteins has been identified in eukaryotes, eubacteria, and archaea (Figure 1; reviewed in Burd and Dreyfuss, 1994; Lewis et al., 1999). Most KH proteins possess more than one copy of the KH motif, with some RNA-binding proteins (e.g., vigilin) containing as many as 15. Nova-1 and Nova-2 both contain three KH motifs.

Mass spectrometry combined with limited proteolysis documented that the three KH motifs of Nova-1 are proteolytically stable domains connected to one another by flexible segments of polypeptide chain (Lewis et al., 1999). Further biophysical characterization of KH domains includes solution NMR structures of KH6 of vigilin (Musco et al., 1996), KH1 of FMR1 (Musco et al., 1997), and KH3 of hnRNP K (Baber et al., 1999), and X-ray structures of Nova-1 and Nova-2 KH3 (Lewis et al., 1999). All five of these compact, globular KH domains are composed of a three-stranded, antiparallel β sheet backed by two or three α helices. There is considerable medical interest in the RNA binding properties of the KH domains of FMR1 and the Nova proteins, because they have been implicated, respectively, in the fragile X mental retardation syndrome and in paraneoplastic opsoclonus-myoclonus ataxia (POMA).

Fragile X syndrome is an X-linked dominant trait with incomplete penetrance that affects ~1 per 1250 males and ~1 per 2000 females, making it the most common form of inherited mental retardation (Nussbaum and Ledbetter, 1995). In most well characterized patients, the syndrome is associated with reduced quantities of fragile X mental retardation protein (FMR1), which contains two KH domains (Siomi et al., 1993b). FMR1 loss of function is typically caused by transcriptional silencing of the *FMR1* gene due to CGG trinucleotide repeat expansion in the 5'-untranslated region (Oberle et al., 1991; Fu et al., 1991; Pieretti et al., 1991; Verkerk et al., 1991), or by significant derangement of the coding segments (Gedeon et al., 1992; Lugenbeel et al., 1995). One particularly severe presentation of fragile X mental retardation is associated with a single base pair mutation in the *FMR1* gene, leading to an Ile→Asn substitution of a highly conserved residue within KH2 (De Boulle et al., 1993).

POMA is a neurodegenerative syndrome, one of several paraneoplastic disorders that arise when systemic malignant tumors express proteins normally sequestered in the central nervous system (Darnell, 1996). The immune system considers these paraneoplastic antigens to be nonself, and the predictable immune response results in neuronal degeneration. In some cases, actual suppression of tumor growth has been observed in patients with paraneoplastic disorders (Darnell and DeAngelis, 1993), making them useful model systems for studying antitumor immunity (Albert et al., 1998; Darnell, 1999). POMA arises primarily in patients with breast/gynecologic tumors that express Nova antigens (Figure 1) (Luque et al., 1991; Buckanovich et al., 1993; Graus et al., 1993; Yang et al., 1998), and is associated with cerebellar (Luque et al., 1991) and brain stem (Hormigo et al., 1994) degeneration.

Considerable effort has been devoted to trying to understand the physiologic roles of the Nova proteins and other paraneoplastic disease antigens, because autoantibody disruption of protein function may contribute to neuronal degeneration (Buckanovich et al., 1996; Okano et al., 1999). Initial RNA binding studies using ribohomopolymers documented that the KH3 domains of Nova-1

[§]To whom correspondence should be addressed (e-mail: burley@rockvax.rockefeller.edu).

^{||}Present address: Department of Biology, Brookhaven National Laboratory, Upton, New York 11973.

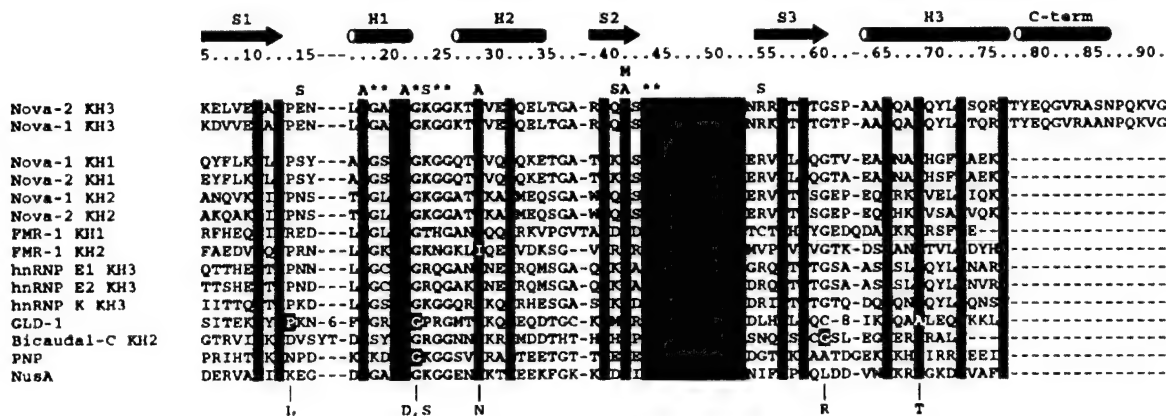


Figure 1. KH Domain Sequence Alignments

Secondary structural elements were assigned on the basis of the X-ray structure. Color coding scheme: yellow, invariant Gly-X-X-Gly motif; red, variable loop; magenta, aliphatic α/β platform. Functional classifications: A, aliphatic stacking interaction; S, side chain-base hydrogen bond, including water-mediated contacts; M, protein backbone-base hydrogen bond; *, van der Waals contact. Sequence alignments of human Nova-1 and Nova-2 KH3, KH1, and KH2, human FMR1 KH1 and KH2 (excluding exon 11), *C. elegans* GLD-1 KH domain, *Drosophila* Bicaudal C KH2, *E. coli* PNP KH domain, and *Methanococcus jannaschii* NusA KH domain. KH domain point mutants discussed in the text are shown in reverse, with the mutant amino acids listed below. Numbers within variable loop segments denote length of the loop.

and Nova-2 are essentially indistinguishable from their full-length counterparts (Buckanovich et al., 1996). In vitro RNA selection experiments (SELEX) (Tuerk and Gold, 1990; Ellington and Szostak, 1992) with full-length and truncated Nova proteins subsequently demonstrated a marked preference for the tetranucleotide 5'-Ura-Cyt-Ade-Y-3', where Y denotes pyrimidine (Buckanovich and Darnell, 1997; Yang et al., 1998; Jensen et al., 2000a). The stem loop crystallization RNA (Figure 2) was identified during the course of exhaustive SELEX experiments (Jensen et al., 2000a) with truncated Nova-1 and Nova-2 proteins (KH3 domain plus additional C-terminal residues; Figure 1). During the past few years, a

number of physiologically relevant RNA targets of the Nova proteins have been proposed (Buckanovich and Darnell, 1997; Jensen et al., 2000b).

Homozygous Nova-1 knockout mice appear to undergo normal gestation and development. Shortly after birth, however, the pups become incapacitated by progressive motor system failure that is correlated with apoptosis of motor neurons in the brainstem and spinal cord, both of which are sites of Nova-1 expression. Death of the homozygous null mice typically occurs within the first two weeks of life, whereas the heterozygotes appear unaffected (Jensen et al., 2000b). These remarkable phenotypic findings occur in the context

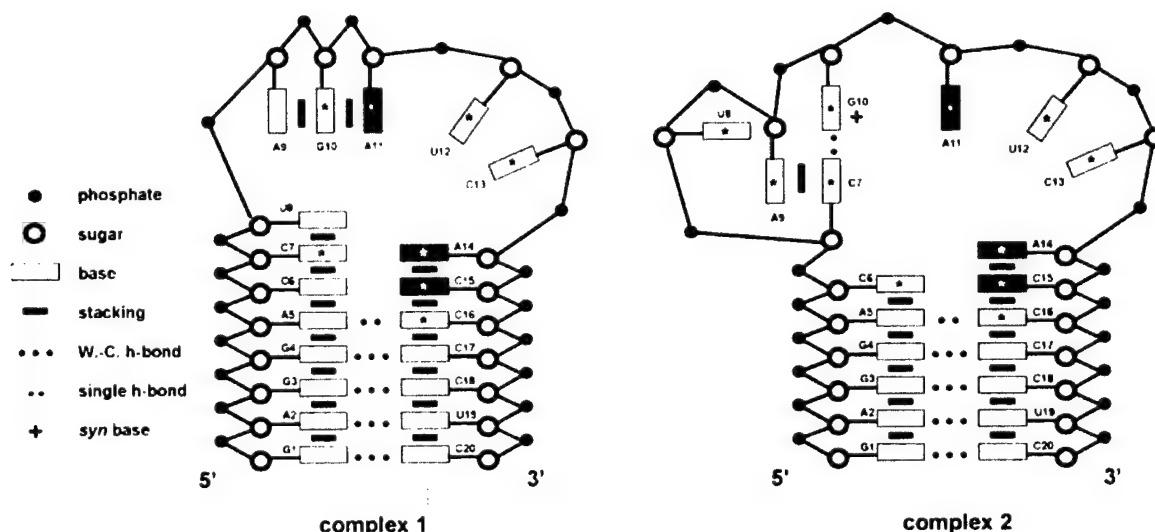


Figure 2. Stem Loop RNA Structures

Schematic drawings of the stem loop crystallization RNAs in complexes 1 and 2. Nucleotides making direct protein contacts are denoted with *. Color coding denotes nucleotides making extensive protein contacts (Ade-11 through Cyt-15; magenta, pink, gold, green, and blue) and the remaining nucleotides (gray).

Table 1. Crystallographic Statistics

Data Set	Resolution (Å)	Reflections, Measured/Unique	Completeness (%) Overall/Outer shell	R _{sym} (%), Overall/Outer Shell
λ ₁ (0.9194 Å)	20.0–2.60	124,023/62,349	96.8/82.4	5.4/20.8
λ ₂ (0.9188 Å)	20.0–2.60	123,214/63,064	97.3/86.8	5.2/19.3
λ ₃ (0.8856 Å)	20.0–2.60	126,070/63,853	98.3/94.9	5.6/23.9
Overall MAD figure of merit = 0.215				
Refinement Data	20.0–2.40	143,241/67,678	98.0/93.7	8.1/21.1
Refinement Statistics	R factor = 0.208, Free R factor = 0.291			
Rms deviations	bond lengths, 0.008 Å bond angles, 1.4° thermal parameters, 3.0 Å ²			

R_{sym} = $\sum |I - \langle I \rangle| / \sum I$, where I = observed intensity, $\langle I \rangle$ = average intensity obtained from multiple observations of symmetry related reflections. Rms bond lengths and rms bond angles are the respective root-mean-square deviations from ideal values. Rms thermal parameters is the root-mean-square deviation between the B values of covalently bonded atomic pairs. Free R factor was calculated with 10% of the data omitted from the structure refinement.

of neuron-specific splicing defects in the glycine and GABA-A receptor pre-mRNAs (Jensen et al., 2000b). The repeated sequence 5'-(Ura-Cyt-Ade-Ura-Y)₃-3' found within the intron immediately upstream of alternatively spliced exon 3A of the glycine receptor α2 pre-mRNA was shown to be a bona fide Nova-binding site (Buckanovich and Darnell, 1997; Jensen et al., 2000b). In vitro splicing studies with transcripts of a glycine receptor α2 mini gene containing the Nova-binding site and exon 3A recapitulated the in vivo observations of (Jensen et al., 2000b), proving that sequence-specific RNA binding by a Nova protein is both necessary and sufficient for regulation of neuron-specific alternative splicing of the glycine receptor α2 pre-mRNA.

In this paper, we report the X-ray structure of a truncated form of Nova-2 bound to a stem loop RNA containing the 5'-Ura-Cyt-Ade-Y-3' core recognition sequence of the glycine receptor α2 pre-mRNA. Our work provides the three-dimensional structure of a KH domain interacting with a high-affinity RNA ligand, demonstrating a novel mechanism by which this RNA-binding motif recognizes single-stranded nucleic acid. The conserved hydrophobic residue implicated in the development of fragile X mental retardation makes an aliphatic stacking interaction with the Ade of 5'-Ura-Cyt-Ade-Y-3', suggesting that the FMR1 protein deficiency phenotype is caused by impairment of RNA binding by KH2 of FMR1. The structure also provides insights into the nucleic acid binding properties of other KH-containing proteins.

Results and Discussion

Crystallization and Structure Determination

The SELEX RNA (Figure 2) plus human Nova-2 KH3 (residues 406–492; Figure 1) yielded high-quality crystals, containing two crystallographically independent copies of a 1:1 protein-RNA complex (denoted 1 and 2) in the asymmetric unit (Experimental Procedures). Experimental phases were obtained using multiwavelength anomalous dispersion (Hendrickson, 1991), with 5-bromo-uracil substituted for Ura-19. A complete structural model for both copies of the protein-RNA complex was built with the aid of noncrystallographic averaging. The current refinement model has a working R factor of 20.8% and a free R value of 29.1% at 2.4 Å resolution, with

excellent stereochemistry (Experimental Procedures, Table 1).

Structural Overview

The crystal structure of the Nova-2 KH3-RNA complex is illustrated schematically in Figure 3. The RNA-bound form of Nova-2 KH3 contains three α helices and three β strands arranged in order S1-H1-H2-S2-S3-H3. Between α helices H1 and H2 there is a Gly-Lys-Gly-Gly 3₁₀-helical portion, which corresponds to the invariant Gly-X-X-Gly motif (yellow) present in virtually all KH domains. β strands S2 and S3 are connected by a 10-residue segment (Lys-43 through Arg-52), referred to as the variable loop (red). Together, the invariant Gly-X-X-Gly motif and the variable loop flank an aliphatic α helix/β strand (α/β) platform that supports RNA recognition, giving the appearance of a molecular vise gripping single-stranded nucleic acid. The only significant structural difference between free (Lewis et al., 1999) and RNA-bound forms of Nova-2 KH3 (root-mean-square deviation [rmsd] between common α-carbons = 0.8 Å) reflects the presence of a C-terminal addition (residues 478–492), which extends α helix H3. The crystallization RNA assumes the predicted hairpin configuration (shown schematically in Figure 2), with electron density visible for all 20 nucleotides in both complexes comprising the asymmetric unit. The stem of the hairpin adopts a standard A-form double-helical conformation, with four Watson-Crick base pairs (Gua-1:Cyt-20, Ade-2:Ura-19, Gua-3:Cyt-18, Gua-4:Cyt-17) and a single hydrogen bond between Ade-5 and Cyt-16 (N1–O2 = 2.4 Å).

Our structure reveals an unusual single-stranded RNA recognition surface, which is neither strictly β sheet nor α-helical as previously suggested by conservation patterns among Nova proteins and the limited structural database of protein-RNA complexes (Lewis et al., 1999). Instead, the extended target RNA (Ade-11, Ura-12, Cyt-13, Ade-14, Cyt-15) lies upon a hydrophobic α/β platform (formed by α helices H1 and H2 and the edge of β strand S2), where it is gripped by the invariant Gly-X-X-Gly motif and the variable loop, both of which are often disordered in structures of free KH domains (reviewed in Lewis et al., 1999). It is remarkable that the platform is exclusively composed of aliphatic side chains, some of which also make van der Waals contacts with conserved residues in the hydrophobic core of the

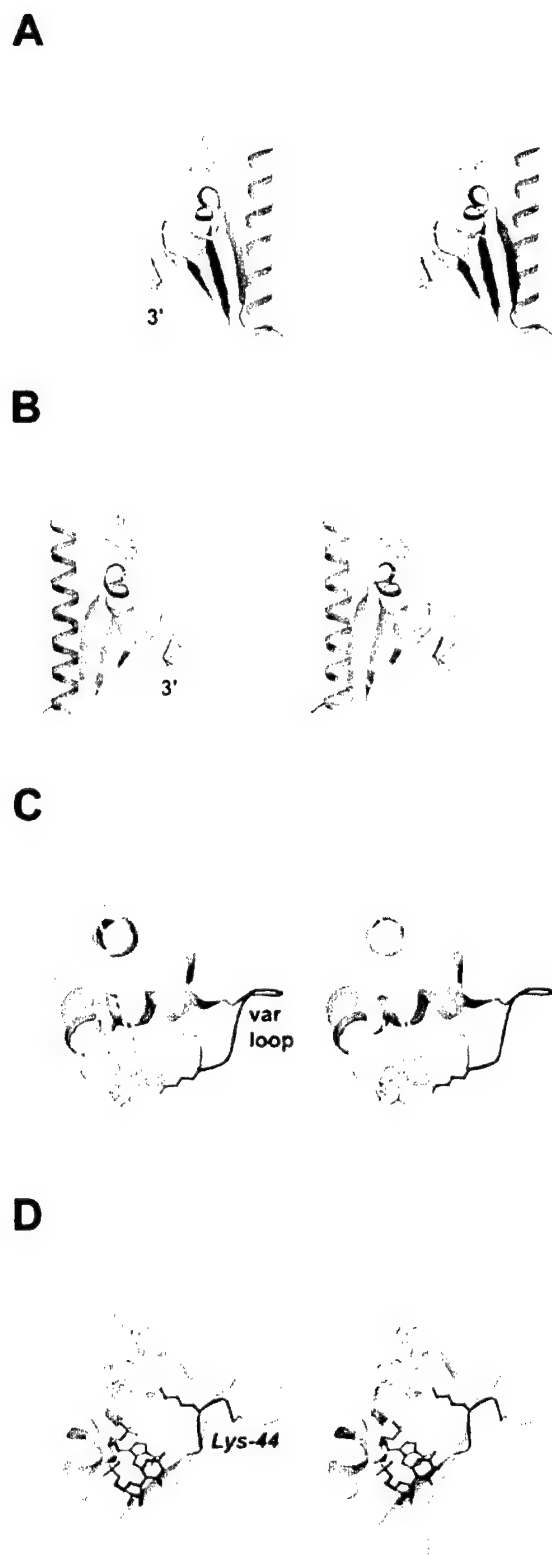


Figure 3. Structure of the Nova-2 KH3-RNA Complex
RIBBONS (Carson, 1991) stereodrawings showing color-coded protein (invariant Gly-X-X-Gly motif, yellow; variable loop, red). Crystallization RNA is included as a color-coded atomic stick figure (Figure

KH domain. Stacking interactions between RNA bases and aromatic side chains typically seen in protein-RNA complexes do not contribute to molecular recognition in this KH domain-RNA complex. Our Nova-2 KH3 co-crystal structure closely resembles that of Nova-1 KH3 bound to the same stem loop RNA, and a detailed comparison will be published elsewhere (H. A. L. et al., unpublished data).

Noncrystallographic Symmetry

The two protein-RNA complexes comprising the crystallographic asymmetric unit (labeled 1 and 2 in Figure 4) are related to one another by a simple 2-fold rotation. With the exception of four nucleotides that adopt an unusual tetraloop conformation in complex 2 but not in complex 1 (Figure 2), the two halves of the asymmetric unit are very similar, effectively ruling out substantial contributions from lattice packing artifacts. α -carbon rmsd is 0.4 Å for protein, and C1' rmsd is 1.2 Å for the crystallization RNA with the tetraloop nucleotides excluded. The structural similarity of the two RNA-protein complexes is even more striking if we limit our analysis to protein α -carbons and C1' atoms of the 5'-Ura-Cyt-Ade-Y-3' core recognition sequence (rmsd = 0.4 Å). A complete analysis of the RNA tetraloop found in complex 2 will be published elsewhere (H. A. L. et al., unpublished data).

The two complexes in the asymmetric unit pack against each other using interactions between the H3 α helices, which are antiparallel with a crossing angle of about 50°. The two stem loop RNAs lie on opposite sides of the noncrystallographic dimer, and extend away from the 2-fold axis. There are no intercomplex protein-RNA contacts. Formation of the asymmetric unit buries 1172 Å² of solvent-accessible surface area, which exceeds the value of 600 Å², the upper limit for adventitious crystal-packing contacts (Janin, 1995). Although it is unclear whether this dimeric organization is of physiologic significance, our previous crystallographic work on the Nova proteins suggests that the dimer depicted in Figure 4 is stable. Our structures of the free forms of Nova-1 and Nova-2 KH3 revealed identical tetrameric arrangements with 22 symmetry (Lewis et al., 1999). One of the two dimer interfaces stabilizing this KH domain tetramer is essentially identical to the binary interactions depicted in Figure 4 (compare Figure 4A with Figure 6 of Lewis et al., 1999).

Recognition of 5'-Ura-Cyt-Ade-Y-3' Core Recognition Sequence of the Glycine Receptor α 2 Pre-mRNA

Overview

Nova-2 KH3 binds to the crystallization RNA using a combination of van der Waals contacts, hydrogen bonds,

2) with a gray ribbon denoting the phosphoribose backbone. α helices are labeled H1, H2, and H3, and β strands are labeled S1, S2, and S3. The N and C termini of the protein, and the 5' and 3' ends of the RNA are labeled. All illustrations and contact distances are derived from complex 2.

(A) View toward the β sheet face.

(B) View toward the α -helical face.

(C) View down the axis of α helix H3, showing the jaws of the molecular vise.

(D) View toward the aliphatic α/β RNA-binding platform, showing the tetranucleotide gripped by the molecular vise.

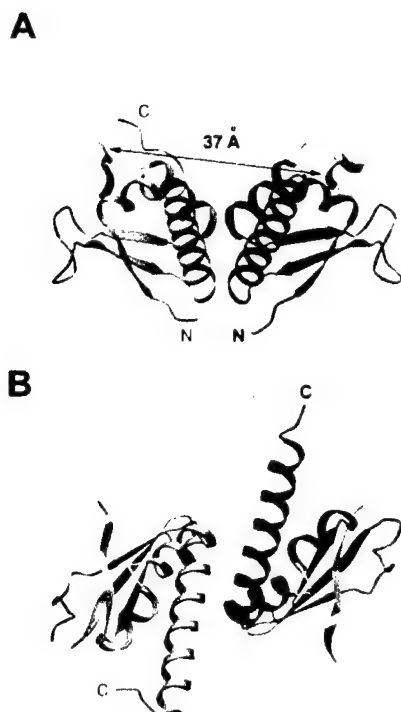


Figure 4. Noncrystallographic Symmetry

(A) Protein/RNA RIBBONS representation of the dimeric assembly comprising the asymmetric unit, viewed perpendicular to the noncrystallographic 2-fold axis. The KH domains have been colored as in Figure 6 of Lewis et al., 1999, and the RNA ribbons are color coded as in Figures 2 and 3.

(B) Viewed along the noncrystallographic 2-fold axis.

and stacking interactions (Figure 5). Although their crystal lattice environments are not the same, the two crystallographically independent protein-RNA complexes in the asymmetric unit make essentially identical protein-ligand interactions, burying portions of both the protein and RNA surfaces (total solvent-accessible area buried = 1900–1967 Å² for each complex). Of the 20 nucleotides in the stem loop RNA, eight interact with protein in complex 1 and 11 in complex 2 (denoted * in Figure 2). The majority of these contacts involve the Ade-11 through Cyt-15 segment (5'-Ade-Ura-Cyt-Ade-Cyt-3'), which is gripped by the jaws of the KH domain molecular vise. Ura-12, Cyt-13, and Ade-14 sit directly atop the α/β aliphatic platform, pinioned between the invariant Gly-X-X-Gly motif and the variable loop. All of the amino acids involved in direct RNA contacts are absolutely conserved between Nova-1 and Nova-2, which explains why these two closely related KH3 domains (92% sequence identity) demonstrate essentially identical nucleic acid binding specificity (Jensen et al., 2000a).

The invariant Gly-X-X-Gly motif (Gly-22, Lys-23, Gly-24, Gly-25 in the KH3 domains of both Nova-1 and Nova-2) serves as one jaw of the vise, and appears to be critical for interactions with the nucleic acid backbone (Figure 5). Gly-22 and Lys-23 make numerous contacts with the sugar-phosphates and bases of Ura-12 and Cyt-13, the first two members of the tetranucleotide

element. Gly-24 and Gly-25 interact with both Cyt-13 and Ade-14. The variable loop makes two RNA contacts in complex 1, with Cyt-7 and Cyt-16. In complex 2, the variable loop interacts with Cyt-6, Cyt-7, Ura-8, Ade-9, Cyt-13, and Cyt-16. Despite their differences in terms of RNA contacts, the variable loops in complexes 1 and 2 are very similar (α -carbon rmsd = 0.6 Å). Not surprisingly, we observed a similar variable loop conformation in our crystallographic study (Lewis et al., 1999) of the free form of Nova-2 KH3 (α -carbon rmsd's = 0.8 Å).

Other residues important for interactions with Ura-12, Cyt-13, and Ade-14 include Glu-14, Val-17, Gly-18, Ala-19, Leu-21, Leu-28, Val-29, Ile-39, Ile-41, Ser-42, and Arg-54. Cyt-15 makes a single contact with Gln-40. In complex 1, the 2' hydroxyl groups of Ura-8, Ade-11, Ura-12, Cyt-13, and Cyt-15 participate in hydrogen bonds with the protein or some other portion of the RNA, making it unlikely that Nova KH3 will bind tightly to a DNA oligonucleotide of the same sequence. In complexes 1 and 2, Ade-11 interacts with both N- (Asn-15) and C-terminal (Arg-83) residues (Figure 5B). This latter contact explains why additional C-terminal residues beyond the confines of both Nova-1 KH3 and Nova-2 KH3 are required for high-affinity RNA binding (Jensen et al., 2000a).

Previously published studies of KH domain mutagenesis are entirely consistent with our Nova-2 KH3-RNA cocystal structure (depicted schematically in Figure 1). Various mutations have been examined in the *C. elegans* GLD-1 single KH domain protein (Jones and Schedl, 1995). Gly-227→Asp and Gly-227→Ser loss-of-function mutations of the first glycine of the invariant Gly-X-X-Gly motif almost certainly interfere with RNA binding via steric interference and/or electrostatic repulsion by an improperly positioned negative charge in the vicinity of the nucleic acid backbone. Similar results were obtained with the KH domain of *E. coli* polynucleotide phosphorylase or PNP (Garcia-Mena et al., 1999). Two other mutations in GLD-1, Gly-248→Arg and Gly-250→Arg, map to the variable loop and cause masculinization of the germline. This portion of Nova-2 KH3 interacts with RNA, thereby suggesting that the *C. elegans* phenotype results from a change in the RNA binding properties of GLD-1. *Drosophila* WHO protein also demonstrates a loss-of-function mutation in the variable loop, Arg-185→Cys, analogous to those seen in GLD-1 (Baehrecke, 1997). In *Drosophila* Bicaudal C (Mahone et al., 1995), a Gly-295→Arg loss-of-function mutation of the conserved Gly at the C terminus of S3 probably disrupts the fold of the KH domain. Similar reasoning explains the deleterious effect of the Ala-294→Thr mutation in GLD-1, which maps to the hydrophobic face of α helix H3.

How Does the KH3 Domain Recognize 5'-Ura-Cyt-Ade-Y-3'?

Ura-12 occurs in the usual anti conformation (Figure 5C), permitting van der Waals interactions with Gly-18 (closest approach = 3.3 Å) and Ala-19 (closest approach = 3.4 Å). Two water molecules form bridges between the base and Lys-23 (N-H₂O = 2.8 Å, H₂O-O2 = 3.4 Å) and Arg-75 (NH₂-H₂O = 3.4 Å, H₂O-N3 = 3.5 Å), providing for indirect detection of Ura at position one of the tetranucleotide consensus sequence. This environment appears to be compatible with Cyt, which was

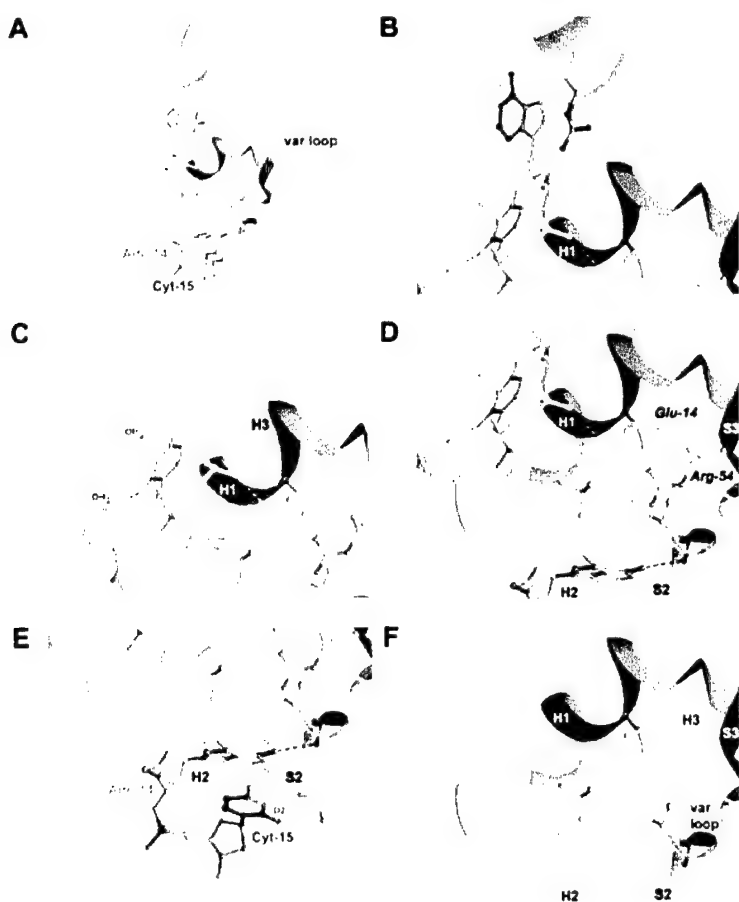


Figure 5. 5'-Ura-Cyt-Ade-Cyt-3' Binding by Nova-2 KH3

RIBBONS drawings showing the RNA-binding surface sandwiched between the invariant Gly-X-X-Gly motif (yellow) and the variable loop (red). The portion of the stem loop crystallization RNA making extensive protein contacts has been included as a color-coded stick figure (Figure 2) with the aliphatic residues (gray) comprising the hydrophobic α/β RNA binding platform. (A) shows the entire complex. (B), (C), and (D) show the environments of Ade-11, Ura-12, and Cyt-13, respectively. (E) shows the environments of Ade-14 and Cyt-15. The side chain of Leu-28 is denoted with *. (F) corresponds to the views used in (D) and (E), with the overlying RNA removed, and shows the aliphatic platform with the jaws of the molecular vise.

tolerated in the context of an exhaustive mutational analysis performed with the crystallization RNA (Jensen et al., 2000a). The larger purine bases would, we believe, be precluded by steric clashes with the protein.

Unlike Ura-12, Cyt-13 was shown to be absolutely required for high-affinity RNA binding (Jensen et al., 2000a). At this position in the tetranucleotide recognition sequence, two protein side chains form a hydrogen bond network that allows the protein to act as a molecular mimic of Gua (Figure 5D). Arg-54 donates two hydrogen bonds to acceptors on Cyt-13 (NH2-O2 = 3.4 Å, NH1-N3 = 3.1 Å), and Glu-14 serves as a hydrogen bond acceptor (OE1-N4 = 2.7 Å). Arg-54 and Glu-14 make an intramolecular interaction (NH1-OE2 = 2.9 Å) that stabilizes their approach to Cyt-13. The location of Cyt-13 is further restricted by an intrastrand hydrogen bond with the phosphate group of Ura-12 (N4-O1P = 3.1 Å), and van der Waals contacts with Val-17 and Leu-21 from α helix H1 (distances of closest approach = 3.2 Å and 3.5 Å, respectively). This complicated network of interactions is only compatible with Cyt at position two of the 5'-Ura-Cyt-Ade-Y-3' core recognition sequence.

Ade-14 is also recognized by a combination of Watson-Crick-like hydrogen bonds with the protein and an intrastrand contact (Figure 5). The backbone of Ile-41 on the edge of β strand S2 mimics Ura (N-N1 = 3.2 Å, O-N6 = 2.7 Å), and O2' of Cyt-13 donates a hydrogen bond that specifies purine (O2'-N7 = 3.5 Å). Ade-14 engages in a number of van der Waals interactions with

the hydrophobic amino acids (Leu-21, Leu-28 and Ile-41) forming the aliphatic α/β platform of the molecular vise (Figure 5E). On the opposite face, an intrastrand stacking interaction with Cyt-15 (distance of closest approach = 3.1 Å) completes the binding site for Ade-14 (Figure 5). This environment is entirely specific for Ade, and is reminiscent of the ATP-binding sites found in the active centers of protein kinases (reviewed in Taylor and Radzio-Andzelm, 1994). Like Nova-2 KH3, these enzymes employ polypeptide backbone mimicry of Ura creating Watson-Crick-like hydrogen bonds with an Ade base sitting atop an aliphatic platform.

The final member of the tetranucleotide core recognition sequence was always found to be a pyrimidine (50% Cyt, 50% Ura) in our Nova SELEX experiments (Jensen et al., 2000a), a finding which is reflected in the environment of Cyt-15 (Figure 5E). Intrastrand stacking interactions are provided by Ade-14 (see above for distance of closest approach) and Cyt-16 (distance of closest approach = 3.3 Å). The observed base contact with Gln-40 is compatible with either Cyt or Ura (NE2-O2 = 3.3 Å). Steric considerations rule out the larger purine bases at this position in the tetranucleotide.

Comparison with Other Protein-RNA Cocrystal Structures

Previously published crystal structures of RNA-binding proteins interacting with single-stranded nucleic acids

include work on RNA recognition motifs or RRM (Oubridge et al., 1994; Price et al., 1998; Deo et al., 1999; Ding et al., 1999; Handa et al., 1999), rho transcription termination factor (Bogden et al., 1999), vaccinia virus methyltransferase VP39 (Hodel et al., 1998), bacteriophage MS2 coat protein (Valegard et al., 1994), trp RNA-binding attenuation protein (Antson et al., 1999), and several tRNA/protein complexes (reviewed in Cusack, 1997). The most common protein structural element employed in RNA recognition is the β sheet as exemplified by the RRM motif, which fueled expectations that RNA binding might occur on the β sheet surface of the KH domain.

In our cocystal structure, single-stranded RNA binds to an α/β aliphatic platform, created by juxtaposition of two α helices and the edge of a β sheet, which is flanked by the invariant Gly-X-X-Gly motif and the variable loop. α/β binding surfaces have been observed in only three other protein-RNA complexes, including VP-39, glutamyl-tRNA synthetase-tRNA^{Glu} (Rould et al., 1991), and threonyl-tRNA synthetase-tRNA^{Thr} (Sakranarayanan et al., 1999). The Nova-2 KH3-RNA complex does not resemble any one of these structures, making KH domains a distinct class of sequence-specific RNA-binding proteins.

Implications for RNA Binding by Full-Length Nova Proteins

Three lines of evidence suggest that 5'-Ura-Cyt-Ade-Y-3' recognition by the Nova proteins is exclusively mediated by KH3. First, all available POMA antisera are directed against Nova-1 and Nova-2 KH3, and incubation of Nova proteins with these antibody mixtures interferes with RNA binding (Buckanovich et al., 1996; Yang et al., 1998). Second, full-length Nova proteins and their KH3 domains display identical RNA binding specificity for 5'-Ura-Cyt-Ade-Y-3', albeit with somewhat reduced affinities for the KH3 domains (Buckanovich and Darnell, 1997; Jensen et al., 2000a). Third, the sequence alignments depicted in Figure 1 argue against 5'-Ura-Cyt-Ade-Y-3' recognition by the remaining KH domains of Nova-1 and Nova-2.

Although Nova-1 KH1 and KH2, and Nova-2 KH1 and KH2 are very similar in amino acid sequence to both Nova-1 KH3 and Nova-2 KH3 (Nova-1: KH1 versus KH3, 39% identity; KH2 versus KH3, 38%; Nova-2: KH1 versus KH3, 36%, KH2 versus KH3, 35%; Figure 1), our cocystal structure of Nova-2 KH3 makes it very unlikely that these four additional Nova KH domains bind tightly to the 5'-Ura-Cyt-Ade-Cyt-3' tetranucleotide. Potentially important differences occur at positions 14 (Ser or Asn instead of Glu), 17 (Ala instead of Val), 19 (Leu or Ser instead of Ala), 21 (Ile instead of Leu), 28 (Ile and Val instead of Leu), 40 (Lys instead of Gln), and 41 (Leu instead of Ile). The chemically significant variation at position 14 (Ser or Asn instead of Glu) almost certainly interferes with recognition of the invariant Cyt, because Glu-14 is critical for molecular mimicry of Gua by the KH domain. Conservative substitutions at positions 21, 28, and 41 are unlikely to affect the ability of the KH domain to act as a molecular mimic of Ura, because polypeptide backbone atoms are used to make hydrogen bonds with N1 and N6 of Ade. Alterations in the aliphatic platform could, however, affect binding of the invariant Ade,

which may be further compromised by changes in the identity of the preceding nucleotide. Similar arguments suggest that the preferences for Ura and pyrimidine at the 5' and 3' ends of the tetranucleotide, respectively, may be altered. Although experimental verification of these predictions is not available for all four Nova KH domains, we know that truncations of both Nova-1 (KH1 plus KH2) and Nova-2 (KH2) do not bind avidly to 5'-Ura-Cyt-Ade-Y-3' (Buckanovich and Darnell, 1997; Jensen et al., 2000a).

What then is the function of the KH1 and KH2 domains of Nova-1 and Nova-2? In the absence of definitive SELEX results, we cannot even be sure that they are capable of high-affinity RNA binding. Further analysis of sequence similarities among Nova KH domains does, however, suggest that they should bind RNA in a similar fashion. All six KH domains of the human Nova antigens share the same Gly-X-X-Gly motif (Gly-Lys-Gly-Gly) and possess highly similar variable loops (Figure 1). Moreover, homology modeling results suggest that they all share the same aliphatic α/β platform between the jaws of the molecular vise (data not shown). Additional SELEX experiments have yielded presumptive evidence that Nova-1 KH1 and Nova-2 KH2 recognize different RNA sequences (K. B. J., K. M., and R. B. D., unpublished observations).

Our cocystal structure demonstrates that individual Nova KH3 domains are not capable of recognizing more than about four or five nucleotides in a row. Gene-specific regulation of pre-mRNA splicing must, therefore, require interplay with additional specificity determinants (i.e., other RNA sequences and/or RNA-bound proteins). The structure of the asymmetric unit depicted in Figure 4 and the presence of three copies of the target tetranucleotide upstream of exon 3A of the glycine receptor $\alpha 2$ pre-mRNA [5'-(Ura-Cyt-Ade-Ura-Y)₃-3'] are consistent with two copies of Nova-1 KH3 binding to the first and third tetranucleotides simultaneously. The distance between the 3' end of the tetranucleotide in complex 1 and the 5' end of the same sequence in complex 2 is about 37 Å, which is compatible with six nucleotides connecting the two recognition sequences. We suggest that the physiologic target of Nova-1 in the glycine receptor $\alpha 2$ pre-mRNA consists of two tetranucleotide sites separated by a six-nucleotide spacer (5'-Ura-Cyt-Ade-Ura-Y-Ura-Cyt-Ade-Ura-Y-Ura-Cyt-Ade-Ura-Y-3', italics denote tetranucleotides interacting with a Nova KH3 dimer). Additional, albeit indirect, support for a Nova dimer interacting with two half-sites comes from work with Nova-1, which documented that simultaneous mutation of Ade→Ura in the first and third tetranucleotide core recognition sequences abolishes RNA binding (Buckanovich and Darnell, 1997). Our dimer model for Nova binding is reminiscent of the behavior of some dimeric transcription factors, such as E2 or the steroid/nuclear receptors (reviewed in Patikoglou and Burley, 1997). This analogy with eukaryotic transcription also underscores the possibility of RNA recognition by Nova-1/Nova-2 heterodimers, which were detected *in vitro* by Yang et al. (1998). Finally, it is remarkable that the RNA binding properties of the Nova proteins resemble those of hnRNP K and hnRNP E1 (Figure 1), which recognize tetranucleotide DICE repeats in the 3' untranslated region of 15-lipoxygenase mRNA (Ostareck et al., 1997).

Implications for Fragile X Syndrome

Our Nova-2 KH3-RNA cocrystal structure also provides some insight into how a single point mutation in KH2 of FMR1 (Ile→Asn, corresponding to Leu-28 illustrated in Figure 5) results in fragile X syndrome mental retardation. Analogous substitutions in various KH domains expressed in *E. coli* yielded denatured proteins, as judged by CD spectroscopy (reviewed in Lewis et al., 1999). Although these observations suggested that FMR1 could be inactivated by a nonconservative substitution destabilizing the structure of KH2, they did not prove that the Ile→Asn point mutation interferes with folding of the KH2 domain in vivo. Feng et al. (1997) used human cell lines derived from a normal volunteer and the fragile X patient possessing the Ile→Asn mutation to demonstrate that mutant FMR1 protein is expressed at normal levels in vivo and remains associated with mRNAs. Further analysis of the mutant protein did, however, reveal a significant functional defect. Unlike wild-type FMR1, which associates with elongating polyribosomes via large mRNP particles, the mutant protein associates with abnormal, smaller RNPs.

The altered Ile in FMR1 KH2 corresponds to Leu-28 in our cocrystal structure (Figure 5E, marked with *). The side chain of this aliphatic residue is directed away from the hydrophobic core of the KH domain, with the CD2 methyl group making van der Waals interactions with Ade-14. Sequence similarities between FMR1 KH2 and the two Nova KH3 domains suggest that they use a common aliphatic α/β RNA-binding platform flanked by the invariant Gly-X-X-Gly motif and the variable loop (Figure 1). Of the four residues comprising this hydrophobic surface, one is identical (Val-17) and the other three are chemically similar to those found in Nova-1 and Nova-2 KH3 (Leu-21 and Leu-28 are both Ile, and Ile-41 is Val). Model building results obtained using our Nova-2 KH3 cocrystal structure suggest that introduction of an isostructural asparagine instead of Leu-28 would leave the hydrophobic interior of the KH domain unchanged, because they both possess CB and CG atoms (Figure 5E). Assuming that the mutant Asn residue adopts the same conformation as Leu-28, OD1 and ND2 of the polar side chain would replace CD1 and CD2 of Leu and change the electrostatic properties of the α/β RNA-binding platform making it significantly less hydrophobic. We believe that such a substitution will affect interactions between FMR1 KH2 and a single-stranded RNA target pinioned between the invariant Gly-X-X-Gly motif and the variable loop. The development of fragile X mental retardation in the patient bearing the Ile→Asn mutation could, therefore, result from perturbation of sequence-specific RNA binding, making it imperative that we fully understand the RNA binding properties of FMR1.

Conclusion

In conclusion, the three-dimensional structure of this unusual protein-nucleic acid complex demonstrates that the KH3 domain of human Nova-2 employs an atypical aliphatic α/β recognition surface that mediates specific, high-affinity binding to 5'-Ura-Cyt-Ade-Y-3'. Detection of a KH domain dimer in the asymmetric unit explains how the Nova proteins can bind to 5'-(Ura-Cyt-Ade-Ura-Y)₃-3' in an intron of the glycine receptor α 2

pre-mRNA and regulate splicing in neurons. Using sequence comparisons with other KH domains, we have been able to propose a general model for interactions of these proteins with RNA. This exercise proved to be particularly pertinent to the fragile X syndrome, because it provides a structural framework with which to try to understand how FMR1 loss-of-function mutations cause the most common form of inherited mental retardation. Our work also provides a starting point for further crystallographic, biochemical, and genetic studies of KH domains and their roles in pre-mRNA splicing, RNA localization, translational control, and mRNA stabilization/degradation.

Experimental Procedures

Protein and RNA Preparation and Crystallization

Nova-2 KH3 (residues 406–492) was expressed in *E. coli* and purified as described in Lewis et al., 1999. The measured molecular mass for this truncated form of Nova-2 was 9752 ± 4 (predicted = 9749), confirming that the protein was neither truncated nor posttranslationally modified during expression or purification. Circular dichroism spectroscopy was used to document that the protein was properly folded, and RNA binding activity measurements were carried out as described in Jensen et al., 2000a. Crystallization RNA was purchased from Dharmacon Research, deprotected according to the manufacturer's instructions, and purified by anion-exchange chromatography followed by desalting.

Nova-2 KH3-RNA cocrystals were obtained by sitting drop vapor diffusion against 25% PEG MME 5000, 150 mM ammonium sulfate, 100 mM MES, pH 6.5, at 20°C with a protein concentration of 155 μ M and a 1:1.25 protein:RNA ratio. After streak seeding, the orthorhombic crystals grow in space group P2₁2₁2₁ ($a = 47.8$ Å, $b = 51.2$ Å, $c = 128.6$ Å), with two protein-RNA complexes/asymmetric unit, and diffract to at least 2.4 Å resolution. For MAD phasing, 5-bromo-uracil was substituted for Ura-19 in the crystallization oligonucleotide.

Data Collection, Structure Determination, and Refinement

Diffraction data at three X-ray wavelengths in the vicinity of the bromine K absorption edge (Table 1) were collected from a single frozen crystal using Beamline X25 of the National Synchrotron Light Source (NSLS) at Brookhaven National Laboratory. The bromine atoms were located using SOLVE (Terwilliger and Berendzen, 1997), and an interpretable electron density map was obtained using MLPHARE and DM (Dodson et al., 1997) with the aid of 2-fold non-crystallographic averaging. After model building, the structure was refined to convergence at 2.4 Å resolution with CNS (Brunger et al., 1998), using diffraction data from CHESS Beamline A1 (Cornell High Energy Synchrotron Source, Cornell University). The final refinement model includes Nova-2 KH3 residues 3–90 (complex 1) and 4–87 (complex 2), both copies of all 20 nucleotides of the crystallization RNA, and 138 water molecules (see Table 1 for refinement statistics). PROCHECK (Laskowski et al., 1993) revealed no unfavorable (ϕ, ψ) combinations in the KH3-RNA complex, with main chain and side chain structural parameters consistently better than average (overall G value = 0.2). Atomic coordinates are available from the Protein Data Bank (PDB Code 1EC6).

Acknowledgments

At the NSLS, we thank Dr. Lonny Berman for help using Beamline X25. At CHESS, we thank Dr. Dan Thiel and the MacCHESS staff for help using Beamline A1. We thank Drs. A. R. Ferré-D'Amaré, J. B. Bonanno, R. J. Buckanovich, R. C. Deo, K. S. Gajiwala, D. Jeruzalmi, J. Kuriyan, J. Marcotrigiano, S. K. Nair, G. A. Petsko, S. Scaringe, N. Sonenberg, and Y. Y. L. Yang for many useful discussions. S. K. B. is an Investigator in the Howard Hughes Medical Institute. This work was supported in part by the Rockefeller University (S. K. B.) and the NIH (R. B. D.). K. M. was supported by the Cornell/Rockefeller/Sloan-Kettering Tri-Institutional M.D.-Ph.D. Pro-

gram and NIH MSTP Grant GM07739. K. B. J. was supported by the Breast Cancer Research Program DAMD17-97-1-7097.

Received October 26, 1999; revised January 4, 2000.

References

- Albert, M.L., Darnell, J.C., Bender, A., Francisco, L.M., Bhardwaj, N., and Darnell, R.B. (1998). Tumor-specific killer cells in paraneoplastic cerebellar degeneration. *Nat. Med.* **4**, 1321-1324.
- Antson, A.A., Dodson, E.J., Dodson, G., Greaves, R.B., Chen, X.-P., and Gollnick, P. (1999). Structure of the trp RNA-binding attenuation protein, TRAP, bound to RNA. *Nature* **401**, 235-242.
- Baber, J.L., Libutti, D., Levens, D., and Tjandra, N. (1999). High precision solution structure of the C-terminal KH domain of heterogeneous nuclear ribonucleoprotein K, a c-myc transcription factor. *J. Mol. Biol.* **289**, 949-962.
- Baehrecke, E.H.D. (1997). who encodes a KH RNA binding protein that functions in muscle development. *Development* **124**, 1323-1332.
- Bogden, C.E., Fass, D., Bergman, N., Nichols, M.D., and Berger, J.M. (1999). The structural basis of terminator recognition by the Rho transcription termination factor. *Mol. Cell* **3**, 486-493.
- Brunker, A.T., Adams, P.D., Clore, G.M., DeLano, W.L., Gros, P., Grosse-Kunstleve, R.W., Jiang, J.-S., Kuszewski, J., Nilges, M., Pannu, N.S., et al. (1998). Crystallography and NMR system: a new software suite for macromolecular structure determination. *Acta Crystallogr. D. Biol. Crystallogr.* **54**, 905-921.
- Buckanovich, R.J., and Darnell, R.B. (1997). The neuronal RNA binding protein Nova-1 recognizes specific RNA targets in vitro and in vivo. *Mol. Cell. Biol.* **17**, 3194-3201.
- Buckanovich, R.J., Posner, J.B., and Darnell, R.B. (1993). Nova, the paraneoplastic Ri antigen, is homologous to an RNA-binding protein and is specifically expressed in the developing motor system. *Neuron* **11**, 657-672.
- Buckanovich, R.J., Yang, Y.Y., and Darnell, R.B. (1996). The onconeural antigen Nova-1 is a neuron-specific RNA-binding protein, the activity of which is inhibited by paraneoplastic antibodies. *J. Neurosci.* **16**, 1114-1122.
- Burd, C.G., and Dreyfuss, G. (1994). Conserved structures and diversity of functions of RNA-binding proteins. *Science* **265**, 615-621.
- Carson, M. (1991). Ribbons 2.0. *J. Appl. Crystallogr.* **24**, 958-961.
- Cusack, S. (1997). Aminoacyl-tRNA synthetases. *Curr. Opin. Struct. Biol.* **7**, 881-889.
- Darnell, R.B. (1996). Onconeural antigens and the paraneoplastic neurologic disorders: at the intersection of cancer, immunity, and the brain. *Proc. Natl. Acad. Sci.* **93**, 4529-4536.
- Darnell, R.B. (1999). The importance of defining the paraneoplastic neurologic disorders. *N. Engl. J. Med.* **340**, 1831-1833.
- Darnell, R.B., and DeAngelis, L.M. (1993). Regression of small-cell lung carcinoma in patients with paraneoplastic neuronal antibodies. *Lancet* **341**, 21-22.
- De Boulle, K., Verkerk, A.J., Reyniers, E., Vits, L., Hendrickx, J., Van Roy, B., Van den Bos, F., de Graaff, E., Oostra, B.A., and Willems, P.J. (1993). A point mutation in the FMR-1 gene associated with fragile X mental retardation. *Nat. Genet.* **3**, 31-35.
- Deo, R.C., Bonanno, J.B., Sonenberg, N., and Burley, S.K. (1999). Recognition of polyadenylate RNA by the poly(A) binding protein. *Cell* **98**, 835-845.
- Ding, J., Hayashi, M.K., Zhang, Y., Manche, L., Krainer, A.R., and Xu, R.M. (1999). Crystal structure of the two-RRM domains of hnRNP A1 (JP1) complexed with single-stranded telomeric DNA. *Genes Dev.* **13**, 1102-1115.
- Dodson, E.J., Winn, M., and Ralph, A. (1997). Collaborative computational project number 8: providing programs for protein crystallography. *Methods Enzymol.* **277**, 620-633.
- Ellington, A.D., and Szostak, J.W. (1992). Selection in vitro of single-stranded DNA molecules that fold into specific ligand-binding structures. *Nature* **355**, 850-852.
- Feng, Y., Absher, D., Eberhart, D.E., Brown, V., Malter, H.E., and Warren, S.T. (1997). FMRP associates with polyribosomes as an mRNP, and the 1304N mutation of severe fragile X syndrome abolishes this association. *Mol. Cell* **7**, 109-118.
- Fu, Y.H., Kuhl, D.P., Pizzuti, A., Pieretti, M., Sutcliffe, J.S., Richards, S., Verkerk, A.J., Holden, J.J., Fenwick, R.G., Jr., and Warren, S.T. (1991). Variation of the CGG repeat at the fragile X site results in genetic instability: resolution of the Sherman paradox. *Cell* **67**, 1047-1058.
- Garcia-Mena, J., Das, A., Sanchez-Trujillo, A., Portier, C., and Montanez, C. (1999). A novel mutation in the KH domain of polynucleotide phosphorylase affects autoregulation and mRNA decay in *Escherichia coli*. *Mol. Microbiol.* **33**, 235-248.
- Gedeon, A.K., Baker, E., Robinson, H., Partington, M.W., Gross, B., Manca, A., Korn, B., Poustka, A., Yu, S., and Sutherland, G.R. (1992). Fragile X syndrome without CCG amplification has an FMR1 deletion. *Nat. Genet.* **1**, 341-344.
- Graus, F., Rowe, G., Fueyo, J., Darnell, R.B., and Dalmau, J. (1993). The neuronal nuclear antigen recognized by the human anti-Ri autoantibody is expressed in central but not peripheral nervous system neurons. *Neurosci. Lett.* **150**, 212-214.
- Handa, N., Nureki, O., Kurimoto, K., Kim, I., Sakamoto, H., Shimura, Y., Muto, Y., and Yokoyama, S. (1999). Structural basis for recognition of the tra mRNA precursor by the Sex-lethal protein. *Nature* **398**, 579-585.
- Hendrickson, W. (1991). Determination of macromolecular structures from anomalous diffraction of synchrotron radiation. *Science* **254**, 51-58.
- Hodel, A.E., Gershon, P.D., and Quirocho, F.A. (1998). Structural basis for sequence-nonspecific recognition of 5'-capped mRNA by a cap-modifying enzyme. *Mol. Cell* **1**, 443-447.
- Hormigo, A., Dalmau, J., Rosenblum, M.K., River, M.E., and Posner, J.B. (1994). Immunological and pathological study of anti-Ri-associated encephalopathy. *Ann. Neurol.* **36**, 896-902.
- Janin, J. (1995). Principles of protein-protein recognition from structure to thermodynamics. *Biochimie* **77**, 497-505.
- Jensen, K.B., Musunuru, K., Lewis, H.A., Burley, S.K., and Darnell, R.B. (2000a). The tetranucleotide UCAY directs the specific recognition of RNA by the Nova KH3 domain. *Proc. Natl. Acad. Sci. USA*, in press.
- Jensen, K.B., Dredge, B.K., Stefani, G., Zhong, R., Buckanovich, R.J., Okano, H.J., Yang, Y.Y.L., and Darnell, R.B. (2000b). Nova-1 regulates neuron-specific alternative splicing and is essential for neuronal viability. *Neuron*, in press.
- Jones, A.R., and Schedl, T. (1995). Mutations in gld-1, a female germ cell-specific tumor suppressor gene in *Caenorhabditis elegans*, affect a conserved domain also found in Src-associated protein Sam68. *Genes Dev.* **9**, 1491-1504.
- Laskowski, R.J., MacArthur, M.W., Moss, D.S., and Thornton, J.M. (1993). PROCHECK: a program to check stereochemical quality of protein structures. *J. Appl. Crystallogr.* **26**, 283-290.
- Lewis, H.A., Chen, H., Edo, C., Buckanovich, R.J., Yang, Y.Y., Musunuru, K., Zhong, R., Darnell, R.B., and Burley, S.K. (1999). Crystal structures of Nova-1 and Nova-2 K-homology RNA-binding domains. *Structure Fold. Des.* **7**, 191-203.
- Lugenbeel, K.A., Peier, A.M., Carson, N.L., Chudley, A.E., and Nelson, D.L. (1995). Intragenic loss of function mutations demonstrate the primary role of FMR1 in fragile X syndrome. *Nat. Genet.* **10**, 483-485.
- Luque, F.A., Furneaux, H.M., Ferziger, R., Rosenblum, M.K., Wray, S.H., Schold, S.C., Jr., Glantz, M.J., Jaekle, K.A., Biran, H., and Lesser, M. (1991). Anti-Ri: an antibody associated with paraneoplastic opsoclonus and breast cancer. *Ann. Neurol.* **29**, 241-251.
- Mahone, M., Saffman, E.E., and Lasko, P.F. (1995). Localized Bicaudal-C RNA encodes a protein containing a KH domain, the RNA binding motif of FMR1. *EMBO J.* **14**, 2043-2055.
- Musco, G., Stier, G., Joseph, C., Castiglione Morelli, M.A., Nilges, M., Gibson, T.J., and Pastore, A. (1996). Three-dimensional structure and stability of the KH domain: molecular insights into the fragile X syndrome. *Cell* **85**, 237-245.

Musco, G., Kharrat, A., Stier, G., Fraternali, F., Gibson, T.J., Nilges, M., and Pastore, A. (1997). The solution structure of the first KH domain of FMR1, the protein responsible for the fragile X syndrome. *Nat. Struct. Biol.* 4, 712-716.

Nagai, K. (1996). RNA-protein complexes. *Curr. Opin. Struct. Biol.* 6, 53-61.

Nussbaum, R.L., and Ledbetter, D.H. (1995). The fragile X syndrome. In *Metabolic Basis of Inherited Disease*, Seventh Edition, C.R. Scriver, A. Beaudet, W.S. Sly, and D. Valle, eds. (New York: McGraw-Hill), pp. 795-810.

Oberle, I., Rousseau, F., Heitz, D., Kretz, C., Devys, D., Hanauer, A., Boue, J., Bertheas, M.F., and Mandel, J.L. (1991). Instability of a 550-base pair DNA segment and abnormal methylation in fragile X syndrome. *Science* 252, 1097-1102.

Okano, H.J., Park, W.Y., Corradi, J.P., and Darnell, R.B. (1999). The cytoplasmic Purkinje antigen *cd2* downregulates Myc function: implications for neuronal and tumor cell survival. *Genes Dev.* 13, 2087-2097.

Ostareck, D.H., Ostareck-Lederer, A., Wilm, M., Thiele, B.J., Mann, M., and Hentze, M.W. (1997). mRNA silencing in erythroid differentiation: hnRNP K and hnRNP E1 regulate 15-lipoxygenase translation from the 3' end. *Cell* 89, 597-606.

Oubridge, C., Ito, N., Evans, P.R., Teo, C.H., and Nagai, K. (1994). Crystal structure at 1.92 Å resolution of the RNA-binding domain of the U1A spliceosomal protein complexed with an RNA hairpin. *Nature* 372, 432-438.

Patikoglou, G., and Burley, S.K. (1997). Eukaryotic transcription factor-DNA complexes. *Annu. Rev. Biophys. Biomol. Struct.* 26, 289-325.

Pieretti, M., Zhang, F.P., Fu, Y.H., Warren, S.T., Oostra, B.A., Caskey, C.T., and Nelson, D.L. (1991). Absence of expression of the FMR-1 gene in fragile X syndrome. *Cell* 66, 817-822.

Price, S.R., Evans, P.R., and Nagai, K. (1998). Crystal structure of the spliceosomal U2B''-U2A' protein complex bound to a fragment of U2 small nuclear RNA. *Nature* 394, 645-650.

Rould, M.A., Perona, J.J., and Steitz, T.A. (1991). Structural basis of anticodon loop recognition by glutamyl-tRNA synthetase. *Nature* 352, 213-218.

Saknaranarayanan, R., Dock-Bregeon, A.-C., Romby, P., Caillet, J., Springer, M., Rees, B., Ehresmann, C., Ehresmann, B., and Moras, D. (1999). The structure of threonyl-tRNA synthetase-tRNA^{Thr} complex enlightens its repressor activity and reveals an essential zinc ion in the active. *Cell* 97, 371-381.

Siomi, H., Matunis, M.J., Michael, W.M., and Dreyfuss, G. (1993a). The pre-mRNA binding K protein contains a novel evolutionarily conserved motif. *Nucleic Acids Res.* 21, 1193-1198.

Siomi, H., Siomi, M.C., Nussbaum, R.L., and Dreyfuss, G. (1993b). The protein product of the fragile X gene, FMR1, has characteristics of an RNA-binding protein. *Cell* 74, 291-298.

Taylor, S.S., and Radzio-Andzelm, E. (1994). Three protein kinase structures define a common motif. *Structure* 2, 345-355.

Terwilliger, T.C., and Berendzen, J. (1997). Bayesian MAD phasing. *Acta Crystallogr.* D53, 571-579.

Tuerk, C., and Gold, G. (1990). Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. *Science* 249, 818-822.

Valegard, K., Murray, J.B., Stockley, P.G., Stonehouse, N.J., and Liljas, L. (1994). Crystal structure of an RNA bacteriophage coat protein-operator complex. *Nature* 371, 623-626.

Verkerk, A.J., Pieretti, M., Sutcliffe, J.S., Fu, Y.H., Kuhl, D.P., Pizzuti, A., Reiner, O., Richards, S., Victoria, M.F., and Zhang, F. (1991). Identification of a gene (FMR-1) containing a CGG repeat coincident with a breakpoint cluster region exhibiting length variation in fragile X syndrome. *Cell* 65, 905-914.

Yang, Y.Y., Yin, G.L., and Darnell, R.B. (1998). The neuronal RNA-binding protein Nova-2 is implicated as the autoantigen targeted in POMA patients with dementia. *Proc. Natl. Acad. Sci. USA* 95, 13254-13259.

Protein Data Bank ID Code

Atomic coordinates for the structure reported in this paper have been deposited in the Protein Data Bank under ID code 1EC6.